

Spring 2015

Persons can speak louder than variables: Person-centered analyses and the prediction of student success

Elisabeth M. Pyburn
James Madison University

Follow this and additional works at: <https://commons.lib.jmu.edu/master201019>



Part of the [Quantitative Psychology Commons](#)

Recommended Citation

Pyburn, Elisabeth M., "Persons can speak louder than variables: Person-centered analyses and the prediction of student success" (2015). *Masters Theses*. 60.
<https://commons.lib.jmu.edu/master201019/60>

This Thesis is brought to you for free and open access by the The Graduate School at JMU Scholarly Commons. It has been accepted for inclusion in Masters Theses by an authorized administrator of JMU Scholarly Commons. For more information, please contact dc_admin@jmu.edu.

Persons Can Speak Louder than Variables:
Person-Centered Analyses and the Prediction of Student Success
Elisabeth M. Pyburn

A thesis submitted to the Graduate Faculty of
JAMES MADISON UNIVERSITY
In
Partial Fulfillment of the Requirements
For the degree of
Master of Arts

Department of Graduate Psychology

May 2015

Acknowledgements

I would first like to thank my advisor, Jeanne Horst. Your selflessness and dedication to helping me with this project (on top of everything else you have to do!) has made this process easy for me. Thanks to you, I feel confident in the quality of my final product and my ability to defend it. You helped me remain calm through panicked emails and uncooperative analyses, and I could not have asked for a more supportive advisor. It has been an absolute pleasure to work with you these past two years, as you have helped me grow academically, professionally, and personally. I look forward to other collaborations in the future!

I would also like to thank my two committee members, Monica Erbacher and Dena Pastor. You have both helped me understand the nuances of mixture modeling and cluster analysis better than I ever thought possible. Your wisdom and insight throughout every step of the process has been invaluable to my learning. Thank you for assisting me in this process!

An additional thank you to my fellow academic cohorts, Heather and Kate. You have both helped me think through my analysis issues that have arisen, and talked me down from the brink of panic when things have gone wrong! I love the opportunities we have had to support each other; I'm so glad we're traveling this road together.

Finally, I must thank my family. Mom and Dad, thank you for always pushing me to excel in school even when I complained about it. Derek, your selflessness and support throughout the past two years (and into a Ph.D. program for the next three!) has made graduate school a breeze for me. I could not have done this without you!

Table of Contents

Acknowledgements.....	ii
List of Tables.....	vii
List of Figures.....	viii
Abstract.....	ix
I. Chapter One: Introduction.....	1
Person-Centered vs. Variable-Centered Approaches.....	1
Classification Analyses.....	3
General Overview.....	3
Usefulness to Psychological Measurement.....	3
Purpose.....	6
II. Chapter Two: Literature Review.....	8
Cluster Analysis.....	9
General Overview.....	9
Initial Considerations.....	10
Impact of outliers.....	11
Transforming data.....	12
Similarity Measures.....	14
Correlational measures.....	14
Distance measures.....	15
Clustering methods.....	17
Hierarchical.....	17
Agglomerative methods.....	17
Divisive methods.....	19
Non-hierarchical.....	20
K-means.....	21
Comparison to hierarchical methods.....	22
Cluster Solution Decisions.....	23
Simple stopping rules.....	23
Complex stopping rules.....	25
Validating clusters.....	26
Summary.....	28
Mixture Modeling.....	28
General Overview.....	28
Initial Considerations.....	30

	Specifying Models.....	30
	Choosing number of classes.....	30
	Estimating parameters.....	31
	Evaluating Model Fit.....	33
	Comparing across models.....	34
	Information criteria (IC)	34
	Why not the chi-square difference test?.....	35
	Likelihood ratio tests.....	35
	Classification-based methods.....	36
	Selecting the final solution.....	37
	Validity Evidence for Classes.....	39
	Comparing Mixture Modeling and Cluster Analysis.....	40
	Main Differences.....	40
	Deciding Between Methods.....	42
	Cluster analysis.....	42
	Mixture modeling.....	44
	Applied Example: Theoretical Background.....	45
	Grouping Variables.....	48
	Goal orientation.....	48
	Work avoidance.....	49
	Help-seeking behavior.....	50
	Validity Evidence Variables.....	53
	Self-acceptance.....	53
	Help-seeking.....	54
	The Big Five.....	54
	Other validity variables.....	56
	Past Research and Present Rationale.....	56
	Research Questions.....	57
III.	Chapter Three: Methods.....	58
	Participants and Procedure.....	58
	Measures.....	59
	Goal orientation.....	59
	Work avoidance.....	59
	Help-seeking	60
	Self-acceptance.....	60
	The Big Five.....	60

	Analysis.....	61
	Data cleaning.....	61
	Cluster analysis.....	61
	Mixture modeling.....	62
IV.	Chapter Four: Results.....	64
	Research Question 1a: Identifying Typologies – Cluster Analysis.....	64
	Analysis.....	64
	Description of clusters.....	65
	Research Question 1b: Validity Evidence – Cluster Analysis.....	65
	Continuous validity variables.....	65
	Categorical validity variables.....	66
	Research Question 1a: Identifying Typologies – Mixture Modeling.....	67
	Analysis.....	67
	Description of classes.....	68
	Research Question 1b: Validity Evidence – Mixture Modeling.....	68
	Continuous validity variables.....	68
	Categorical validity variables.....	69
	Research Question 2: Differences between Profiles.....	70
	Research Question 3: Predicting GPAs with Profiles.....	71
	Non-nested regression models.....	73
	Nested regression models.....	73
	Cluster/Class 3 as comparison group.....	73
	Cluster/Class 2 as comparison group.....	75
	Cohen’s <i>d</i> comparisons.....	75
V.	Chapter Five: Discussion.....	77
	Brief Overview.....	77
	Research questions.....	77
	Variables of interest.....	77
	Qualitative Distinction of Profiles: Cluster Analysis.....	78
	Interpretation of clusters.....	78
	Validity evidence.....	79
	Conclusions.....	82
	Qualitative Distinction of Profiles: Mixture Modeling.....	83
	Interpretation of classes.....	83
	Validity evidence.....	84
	Conclusions.....	85

What Do These Profiles Reveal?	86
Differences between cluster analysis and mixture modeling.....	86
Final solution differences.....	86
Validity evidence.....	87
So which is “better” – mixture modeling of cluster analysis?.....	88
Student success.....	89
Implications, Limitations, and Future Research.....	90
Conclusion.....	93
Tables.....	94
Figures.....	105
Appendices.....	108
References.....	109

List of Tables

Table 1. Example of Using Agglomeration Coefficients as a Stopping Rule.....	94
Table 2. Demographic Information for Participants.....	94
Table 3. Chi-square Results: Gender by Major.....	95
Table 4. Subscale Means and Intercorrelations: Classification and Validity Variables.....	96
Table 5. Agglomeration Coefficients - Last 10.....	97
Table 6. Means and SDs of Final Clustering Solution.....	97
Table 7. ANOVA Results for Continuous Validity Variables (Clusters)	98
Table 8. Chi-square Results: Cluster (Cluster Analysis) and Class (Mixture Modeling) by Major.....	99
Table 9. Fit Indices for the Three Mixture Model Parameterizations.....	100
Table 10. Class Means by Classification and Validity (Auxiliary) Variables.....	101
Table 11. Covariances and Variances by Class.....	102
Table 12. Classification Table: Cluster by Class.....	102
Table 13. Regression Values for the Prediction of Spring GPA from Cluster and Class (Cluster/Class 3 as Comparison Group)	103
Table 14. Regression Values for the Prediction of Spring GPA from Cluster and Class (Cluster/Class 2 as Comparison Group)	104
Table 15. Cohen's d Comparison of GPA Means across Classes (by Assignment Type) and Clusters.....	104

List of Figures

Figure 1. Illustration of how structure can be imposed on data where no structure exists.....	105
Figure 2. Illustration of the issues with using correlation as a measure of similarity.....	105
Figure 3. Visual representation of the concept of Euclidean distance.....	105
Figure 4. Possible student profiles resulting from cluster analysis or mixture modeling, utilizing the variables of study.....	106
Figure 5. Z-score means by cluster for the three-cluster hierarchical agglomerative cluster analysis solution.....	106
Figure 6. Z-score means by cluster for the final three-cluster <i>k</i> -means cluster analysis solution.....	107
Figure 7. Z-score means by class for the final three-class mixture modeling solution (modal assignment).....	107

Abstract

In order to ensure that analyses are appropriate for one's research question(s), it is important to consider whether a person-centered or variable-centered approach is needed. Person-centered approaches are often not considered in situations for which they would be appropriate. To that end, a description of the characteristics and procedures of two common person-centered analyses (cluster analysis and mixture modeling) are provided. Although both analyses accomplish the same general aim – to group persons based on their similarity on a series of variables, thus providing ease of interpretation – the methods employed for each analysis differ considerably. As illustration, both analyses were applied to a sample of student data. Scores on six measures, collected during a university-wide assessment day, were used to group students via cluster analysis and mixture modeling – mastery approach, performance approach, and performance avoidance goal orientations; work avoidance; and two help-seeking orientations. Profiles were then compared to identify similarities and differences between analysis solutions. Predictive utility of the profiles was also assessed by entering them into a regression predicting GPA.

Both analyses resulted in three groups for their final solutions, based on decision criteria considered best practice for each analysis. Groupings were supported by validity evidence. Patterns of means between the cluster analysis and mixture modeling profiles were similar in terms of overall ranking and cluster-to-class assignment; however, qualitative differences among the profiles were also identified. Specifically, the mixture modeling classes did not differ very much on work avoidance and the two help-seeking variables, whereas the cluster analysis classes did. Cluster and class sizes were also

discrepant, with Class 3 consisting of many more students than any of the other clusters or classes. Regression analyses indicated that neither the clusters nor the classes meaningfully predicted GPA.

Researchers should consider person-centered analyses if their research questions so dictate; however, the different processes employed in mixture modeling and cluster analysis require that researchers also consider which analysis is most appropriate for their needs. Prior hypotheses regarding population and/or sample structure should also be considered.

CHAPTER ONE

Introduction

In a special edition of *Contemporary Educational Psychology*, Marsh and Hau (2007) put forth a serious issue facing educational and psychological research. They posited that far too many substantive researchers fail to practice good methodology, while methodologically-oriented researchers fail to perform research that is of interest to those involved with substantive domains. Their solution to this problem was the concept of methodological synergy – a fusion of substantive research with sound methodological practices. The mismatch between substantively interesting and methodologically sound research may stem from several deep-running problems that plague today's social science research community; however, an awareness of the importance of methodological synergy can help raise the quality of research being conducted. One fundamental consideration when attempting to develop methodologically synergistic research involves the orientation one will take: does the research question dictate a variable-oriented or person-oriented approach?

Person-Centered vs. Variable-Centered Approaches

The majority of univariate and multivariate statistical analyses employed in psychological research is variable-centered – that is, hypotheses and research questions are typically framed in terms of the variables and their relationship to or predictive ability for the outcome of study (Bergman & Magnusson, 1997; Laursen & Hoff, 2006). However, in recent decades, there has been a push – especially among developmental researchers (e.g., Bergman & Magnusson, 1997) – to also consider a person-centered approach to some research questions. Although variable-centered analyses are certainly

appropriate when seeking predictors of an outcome, they are not necessarily appropriate when seeking to make statements about individuals (Bergman & Magnusson, 1997). This is because variable-centered methods are focused on the structure of the variables *across* persons, rather than the patterns of responding *within* persons (Marsh, Lüdtke, Trautwein, & Morin, 2009). An additional assumption underlying variable-centered methods is that the variable/outcome relationship is the same across all members of the population; however, this is often not the case (Laursen & Hoff, 2006).

In contrast, the person-centered approach permits examination of the patterns and relationships among the variables at the level of the individual. Whereas the assumption underlying the variable-centered approach is that there is population homogeneity in regards to the variable/outcome relationship, an assumption of heterogeneity underlies the person-centered approach – that is, different patterns of relationships occur for different people (Bergman & Magnusson, 1997; Laursen & Hoff, 2006). Person-centered methods provide a more comprehensive and holistic view of the persons being studied, as well as a more realistic understanding of the multivariate outcomes (i.e., patterns of responses) than variable-centered methods (Magnusson, 1998).

It is important to note that both person- and variable-centered methods can be employed together when appropriate. Each approach provides a different perspective on the data, and these perspectives can be effectively joined to create a more complete picture of the results (Hair et al., 1998). For example, the variable patterns observed via person-centered methodology can be used as variables themselves in variable-centered techniques. This fusion of methodology provides an overarching picture that can make complex relationships more readily apparent (Hair et al., 1998; Laursen & Hoff, 2006).

Classification Analyses

General Overview

Logically, person-centered research questions should be answered by using analytic methodology that is also person-centered. It is here that classification analyses – also called taxonometric methods (e.g., MacCallum, Zhang, Preacher, & Rucker, 2002) – come into play. Classification analyses group persons based on their similarity on certain variables of interest (Milligan & Hirtle, 2012), shifting the focus from the variables to the person. Historically, classification analyses were more commonly employed in psychiatry rather than psychology due to the medical necessity of categorizing patients according to diagnoses (Bergman & Magnusson, 1997). However, with the recent advent of powerful computers (Magidson & Vermunt, 2002) as well as increased focus on person-centered methodology (e.g., Bergman & Magnusson 1997; Magnusson, 1998; von Eye & Bogat, 2006), classification analyses are seeing increased usage in the psychology research community at large (Bergman & Magnusson, 1997). Two common classification-type person-centered analyses that will be the main focus of this paper are cluster analysis and mixture modeling (Magnusson, 1998).

Usefulness to Psychological Measurement

Some statisticians and research methodologists object to the use of classification techniques like cluster analysis or mixture modeling altogether. MacCallum et al.'s (2002) well-known article criticizing the practice of dichotomizing continuous variables cautions against utilizing classification analyses unless absolutely necessary, positing that groups identified by such techniques are “probably an oversimplification and potentially misleading” (MacCallum et al., 2002, p. 34). However, not all methodologists feel the

same way (e.g., Bauer & Shanahan, 2007; Bergman & Magnusson, 1997; Marsh et al., 2009). Classification techniques can in fact be an effective and understandable way to capture complex interactions in data with many predictor variables (Bauer & Shanahan, 2007). The number of interactions requiring interpretation in regression analyses, for example, increases exponentially with each predictor added. Classification analyses capture these patterns and relationships in a parsimonious way, allowing for easier interpretation and understanding (Bauer & Shanahan, 2007).

The use of classification analyses also provides an empirically-based way for researchers and the general public alike to meaningfully conceptualize information. Human beings are naturally inclined to group objects based on common characteristics in ways that make them easier to remember and understand (Tan, Steinbach, & Kumar, 2006); classification analyses can provide empirical support for such groupings. In the same vein, the solutions that arise from classification analyses can support or be supported by classes or clusters already theorized to exist in certain populations or samples. Although the clusters themselves must be interpreted cautiously on their own, generating already-theorized groups can help lend support to the theory (Hair et al., 1998).

In addition to providing a way to parsimoniously conceptualize data, the usefulness of classification analyses to psychological measurement can be seen in the difference between variable-centered and person-centered approaches to psychological research. As mentioned previously, the person-centered approach considers the individual holistically, echoing Gestalt psychology's assertion that the whole is more than the sum of its parts (Magnusson, 1998). Person-oriented theorists believe that the complexity of a

person's psychological functioning cannot be properly understood by examining individual variables in isolation from other variables that might also impact a person's psychological functioning. The need to study the individual holistically can be best understood when considering longitudinal research, in which the focus is on patterns across time. Participants in a longitudinal study may differ from one another on levels of a particular individual variable at a given time; but at the person level, of more interest is how participants change differently across time. That is, the focus is on *patterns* of individual responses over time, rather than the *variables* in isolation. Moreover, the person-centered longitudinal researcher is interested in the holistic functioning of the individual, which is represented in the interaction of the variables across and with time to form differing patterns of change (Magnusson, 1998; Marsh et al., 2009).

Although it is easy to see how the person-oriented approach applies to longitudinal studies, it is also applicable to most, if not all, multivariate psychological research. Arguably, the purpose of psychological research is to understand the cognitive and behavioral functioning of persons (Magnusson, 1998). However, the variable-centered approach, with its traditional focus on variables and their relationships to each other and the criteria, treats *variables* as if they are the actors rather than the *person* (Coleman, 1986). Researchers who use the variable-centered approach assume that interrelationships among variables are the same for all persons being studied. However, this is often not the case. In all research, it is important to ensure that one's statistical approach appropriately matches the model of study (Wilkinson, 1999). It thus makes much more sense to examine the patterns of relationships – i.e., to take the person-oriented approach – than to focus on the variables alone (Magnusson, 1998).

The person-centered perspective has implications for psychological measurement, as it requires a shift in the understanding of what an individual's "score" on an instrument means. The variable-oriented approach examines the score in relation to other people's scores on the same scale. In contrast, the person-oriented approach examines the score in relation to the same person's scores on the other instruments – that is, how each score fits into the multivariate pattern of all scores across the individual. A score is only understandable when considered in context (Magnusson, 1998).

If there are different patterns across persons, then it logically follows that some individuals' patterns will be more similar than others, and can and should be grouped together to facilitate understanding. It is here that grouping techniques such as cluster analysis and mixture modeling become invaluable tools for the multivariate researcher. These groupings can be used as variables in other analyses, providing a more complete picture of how the factors of study influence the individual than the variables alone would be able to do (Bauer & Shanahan, 2007; Magnusson, 1998).

Purpose

Given the importance of utilizing person-centered analyses for person-centered research questions, it is vital that researchers are aware of what analyses exist and how to conduct them. To that end, this paper will provide a detailed description and comparison of two useful person-centered analyses – cluster analysis and mixture modeling. To do so, the methodological literature pertaining to each technique will be examined, points of disagreement among analysts will be discussed, and comparisons between the two analyses will be made. Additionally, situations in which one analysis may be more

appropriate than another will be described in an effort to assist researchers in making a decision about which technique to use.

In the spirit of methodological synergy, the two techniques will also be used to analyze an actual dataset. An applied example will provide the opportunity for a concrete explanation of the nuances of each technique, while issues that arise with the data will allow the demonstration of different ways of addressing problems in practice. In sum, the purpose is to inform the reader about not only the value of person-centered approaches to research, but also empirically-based methods of exploring person-centered research questions.

CHAPTER TWO

Literature Review

Given the field of psychology's focus on the individual, a person-centered approach to research clearly has a place in psychological studies. Despite this fact, many methodologists and researchers continue to utilize variable-centered methods in situations where person-centered methods would be more appropriate (Bergman & Magnusson, 1997). Perhaps this is because many researchers are unaware of the important distinctions between the two methodological approaches; or, if they are aware, perhaps they are unsure of what analytical tools are available to conduct person-centered research. Although the overwhelming prevalence of variable-centered research makes this lack of knowledge understandable, it is important for psychological researchers to be aware of methodology appropriate for all types of research questions (Laursen & Hoff, 2006). Such awareness ensures that research is being conducted appropriately and in a manner that will provide the most insight into the object of study.

Two popular person-centered analyses are cluster analysis and mixture modeling. Both of these methods can be grouped under the heading of classification analyses – that is, analyses that group objects (typically people in psychological settings) based on similarity. Such groupings permit the individual to be examined holistically, across a range of variables. Thus, classification analyses are considered person-centered in that they are focused on the person as a whole rather than individual variables. Cluster analysis and mixture modeling have many applications in psychological research, from educational psychology to developmental psychology to psychological measurement – basically any scenario in which the person is the primary object of interest. It is thus

important for researchers to understand how to conduct these analyses and in what research situations they are most applicable.

Cluster Analysis

General Overview

One popular person-centered method is a multivariate technique called cluster analysis. The primary purpose of cluster analysis is to create groups of objects (which in the case of most social science research means people) based on certain common characteristics. These characteristics are defined by a set of variables known as the cluster variate; the variables in the cluster variate could include demographics (age, race, gender, etc.), scores on a set of measures, or levels of a latent variable. Unlike most other multivariate analyses, the purpose of cluster analysis is not to estimate the variate; rather, the purpose is to use the researcher-defined variate to compare objects (Hair et al., 1998). These objects (i.e. people) are grouped in such a way as to maximize within-group homogeneity and between-group heterogeneity – that is, objects within a group should be similar to each other, based on the variables in the cluster variate, but dissimilar to objects in other clusters (Milligan & Hirtle, 2012; Pastor, 2010).

Although cluster analysis is a multivariate technique, it is unlike many other multivariate techniques in that the groups are not known prior to beginning the analysis. Discriminant analysis, for example, seeks to differentiate among *known* groups based on a set, or composite, made up of the same type of variables that would be included in the cluster variate. However, where the intent of discriminant analysis is to examine multivariate differences in *known* groups (e.g., gender), the primary purpose of using cluster analysis is to identify groups, based on the variables (Pastor, 2010). Because of

this, the clusters are wholly dependent not only on the variables chosen by the researcher to make up the cluster variate, but also on the sample itself. Additionally, cluster analysis is strictly exploratory and non-inferential; because it is designed to impose a grouping structure on the data, it will do so whether or not groups actually exist in the data. To illustrate this, see Figures 1a and 1b (adapted from Everitt, Landau, Leese, & Stahl, 2011). Figure 1a displays a set of data points (representing persons) that clearly have no inherent structure or groupings. However, a researcher could request four clusters when applying cluster analysis to this dataset, and would probably get a grouping division something like Figure 1b, in which each “quadrant” represents a cluster. Although the divisions in this figure are clustering the most similar persons together, dividing the data in this way is meaningless and potentially misleading. It is for this reason that it is important for cluster analysis researchers to choose the cluster variate carefully, to ensure that their samples are representative of the population, and to engage in further analysis beyond just creating groups (Hair et al., 1998; Pastor, 2010).

Initial Considerations

Two of the most important initial steps when conducting a cluster analysis are the identification of 1. the objects to be classified and the population from which they will be drawn and 2. the variables that will make up the cluster variate (Lorr, 1983; Milligan & Hirtle, 2012). The objects, as the focus of the study, are the primary basis of the analysis. However, equally important are the variables, because the cluster solution is based solely on the objects’ values on the variables (Milligan & Hirtle, 2012; Pastor, Barron, Miller, & Davis, 2007). The cluster solution may differ dramatically depending on which variables are selected, so it is important for the researcher to identify the appropriate variables prior

to beginning analysis. The selection of variables may be based on practical or theoretical considerations (or both), but researchers should have an adequate rationale for their choice and should clearly outline this rationale when writing about their findings (Hair et al., 1998; Pastor, 2010). It is also important to not include too many irrelevant variables – that is, variables that do not have a bearing on identifying the clusters. Irrelevant variables may “mask” the true cluster structure and lead to a misleading solution (Milligan, 1980). Once the decision about the objects and the variables has been made, the researcher can move on to other steps in the research process.

Impact of outliers. Although variable selection is extremely important to the eventual clustering solution, researchers should also carefully examine the objects (i.e., cases) in their sample. Of particular importance is examining data for outliers, which can unduly influence results in potentially unfavorable ways. Whether outliers are the result of a genuinely unusual case, an instance of an underrepresented group in the population, or a data error, they can cause the cluster solution to be unrepresentative of the true structure inherent in the population (Pastor, 2010). However, the impact of outliers on the final clustering solution may depend on the type of clustering method used. One simulation study found that hierarchical methods in particular (both hierarchical and non-hierarchical methods will be described in detail later in this paper) tend to be markedly negatively affected by outliers. In contrast, the non-hierarchical centroid method was almost unaffected by outliers. In data with a large number of outliers, then, it may be advisable to utilize a non-hierarchical method rather than a hierarchical one (Milligan, 1980; Milligan & Hirtle, 2012).

There are other ways of dealing with outliers, however. As with most analyses, the outlying cases could simply be deleted. Alternatively, cluster analysis could be conducted both with and without the outliers included, and the clusters examined to determine whether the outliers are unduly affecting the solution (Milligan & Hirtle, 2012). Whatever method is chosen, it is important to report and justify one's reasons for doing so (Pastor, 2010).

Transforming data. The similarity measures used to generate the clusters – and thus the clustering solutions themselves – may be substantially impacted when the variables in the cluster variate are on different scales (Fleiss & Zubin, 1969). This is because the variable(s) with the largest standard deviations tend to have the most impact, in effect weighting the clustering solution to be biased towards such variables (Anderberg, 1973). One popular method of correcting for this is to standardize the variables (i.e., convert them to z-scores; Fleiss & Zubin, 1969). Standardization has several advantages beyond the fact that it corrects for unequal weighting in the cluster solution. It makes it easier to compare among the variables, and also allows the researcher to change the scale (e.g., from hours to minutes) without affecting the standardized value (Hair et al., 1998). However, a z-score transformation is not the only method of standardization, nor is it necessarily the best method (Milligan, 1996; Milligan & Cooper, 1988; Milligan & Hirtle, 2012; Steinley, 2004).

One issue with using z-score transformations to standardize variables involves which standard deviations are used for the transformation. In the case of cluster analysis, the within-group standard deviations are seldom, if ever, known (Milligan & Cooper, 1988). As a result, the overall sample standard deviation is used instead. However, doing

so often “dilutes” the cluster separation, causing less pronounced differences in some cases and more pronounced differences between members of the same cluster in others (Fleiss & Zubin, 1969). Thus, some researchers strongly advise against using z-score transformation in many cases (e.g., Milligan & Cooper, 1988; Milligan & Hirtle, 2012). These researchers argue that standardizing variables would be inappropriate in cases where theory dictates that the clusters exist in the untransformed variable space (Milligan, 1980). In these cases, standardizing the variables by z-score conversion may cause the true solution to be distorted. As a result, it is advisable to consider other methods of standardizing variables (Fleiss & Zubin, 1969; Milligan & Hirtle, 2012).

Milligan and Cooper’s (1988) simulation study tested several different standardization methods for accuracy. Most of the methods they tested do not use the standard deviation, thus avoiding the problem described in the previous paragraph. The most effective standardization techniques utilized the range of the variable in the denominator:

$$\frac{x}{\text{Max}(x) - \text{Min}(x)}$$

and

$$\frac{x - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)}$$

These two standardization methods performed consistently well across the four clustering methods examined by Milligan and Cooper (1988). The superiority of range-based standardization methods has also been borne out in subsequent studies, and should thus be seriously considered as an alternative to z-score conversion methods (Milligan & Hirtle, 2012; Steinley, 2004).

Similarity Measures

In order to group objects into clusters – the primary purpose of cluster analysis – the criteria for determining similarity among objects must first be decided upon. This criterion can then be used to group the most similar objects together. Although similarity seems like a relatively simple concept, there are in fact several different ways in which it can be determined (Everitt et al., 2011; Fleiss & Zubin, 1969; Milligan & Cooper, 1987).

Correlational measures. One similarity method that has seen some historical use involves correlating every pair of objects' values for each variable, to produce a correlation coefficient matrix. This matrix is then used in a Q-type factor analysis, and the resulting factors are considered the clusters. Each object is assigned to the factor/cluster on which it loads most strongly. Although this method may make logical sense, there are several problems with using correlations as the measure of similarity and subsequently following the correlations with factor analysis. First, an observed high correlation between two variable patterns (or profiles) could occur if the profiles were parallel yet far apart in terms of magnitude. Second, the profiles need not even be parallel to have a high correlation as long as they are linearly related. That is, they could have a high correlation, but not be practically similar (Fleiss & Zubin, 1969; Hair et al., 1998). Figure 2, adapted from Fleiss and Zubin (1969, p. 237), illustrates this second point. Test-taker 2's scores are exactly twice the scores of test-taker 1, plus one (e.g., for Test A, test-taker 1 received a (-1). $(-1) + (-1) = (-2)$, and $(-2) + (1) = (-1)$, which is test-taker 2's score for Test A). Despite the clear dissimilarity of these two score profiles, the correlation between test-taker 1 and 2 is a perfect +1. Further complicating matters is test-taker 3, whose scores are identical to test-taker 1 except for the score on test E. From

a practical standpoint, test-taker 3 is most similar to test-taker one. However, the correlation between 1 and 3 is .99 – lower (albeit only slightly) than the correlation between the more dissimilar test-takers 1 and 2! Clearly, using correlation as a measure of similarity poses problems in cluster analysis.

Distance measures. Technically, distance measures are a measure of *dissimilarity* rather than similarity (Milligan & Cooper, 1987). They involve theoretically plotting each object in multidimensional space, with as many dimensions as there are variables. The larger the “distance” between the points is, the more dissimilar the objects are (Everitt et al., 2011). Logically, objects that are closest together in this multidimensional space are grouped together to form the clusters (Fleiss & Zubin, 1969; Hair et al., 1998). There are many types of distance measures for all different kinds of data (i.e., continuous, categorical, or nominal); however, this paper will only address two of the most common, which are used for continuous data (Everitt et al., 2011). The interested reader is referred to Anderberg, 1973; Everitt et al., 2011; and Lorr, 1983 for a more comprehensive list of available similarity measures.

Euclidean distance is the most common of all the distance measures (Everitt et al., 2011), and is obtained by calculating the hypotenuse of a right triangle formed from the two points of interest (see Figure 3, adapted from Hair et al., 1998, p. 486). Euclidean distance is intuitively appealing, as it is representative of the actual physical distance between two points, as can be seen in the formula:

$$distance = \left[\sum_{k=1}^p w_k^2 (x_{ik} - x_{jk})^2 \right]^{1/2}$$

where x_{ik} and x_{jk} are the values of the k th variable for persons i and j (Everitt, 2011). w_k is a weighting term that can be applied to the variable, but is often set to 1 (though it does not have to be; Everitt, 2011; Milligan & Cooper, 1987). Squared Euclidean distance is often used to avoid having to take the square root of the calculated distance (Hair et al., 1998).

Another commonly used distance measure is the city-block method, which is similar to Euclidean distance. City-block distance is sometimes also called taxicab or Manhattan distance, since it measures distance by using a grid system resembling city blocks to determine the shortest path between the two points (Everitt et al., 2011; Milligan & Cooper, 1987). Whereas the Euclidean distance measure uses the squared difference between 2 points, the city-block method uses the absolute value of the difference:

$$distance = \sum_{k=1}^p w_k |x_{ik} - x_{jk}|$$

where, once again, x_{ik} and x_{jk} are the values of the k th variable for persons i and j and w_k is the weighting term (Everitt et al., 2011).

Choosing the correct distance measure is extremely important, as there is evidence that choosing incorrectly may lead to incorrect cluster solutions (Milligan & Cooper, 1987). As mentioned previously, the Euclidean and city-block distances are to be used with continuous variables (Everitt et al., 2011); however, data may also be categorical or nominal. When data are not continuous, it would be best to use a more appropriate similarity measure (e.g., chi-square based measures; Anderberg, 1973). One should also consider the clustering method that will be used, as some methods work best

with certain similarity measures. It is thus important to be aware of issues and past research prior to choosing a similarity measure (Everitt et al., 2011; Milligan, 1996).

Clustering Methods

Although the overarching purpose of cluster analysis is to create homogenous groups, there are several different ways to go about the actual clustering process. These methods, or clustering algorithms, can be broken down into two main categories: hierarchical and non-hierarchical. Different methods will likely result in different clustering solutions, so it is important to understand them prior to selecting a method (Hair et al., 1998).

Hierarchical. Hierarchical clustering methods take one of two forms. In the agglomerative method, each case begins the process in its own cluster (i.e., initially there are the same number of clusters as there are objects). Clusters are then combined, one by one, with nearby clusters until all clusters/cases have been joined into one large cluster. In contrast, the divisive method works in reverse, with all cases grouped together in a single cluster and gradually split off to make smaller clusters. Although the procedures are essentially mirror images of one another, agglomerative methods are the ones typically used in statistical software packages as well as in most research employing cluster analysis (Hair et al., 1998; Johnson, 1967; Milligan & Cooper, 1987; Milligan & Hirtle, 2012).

Agglomerative methods. The main difference among agglomerative algorithms is the way in which similarity is calculated. Because the clusters that are to be combined are determined by how similar (or, in some cases, dissimilar) they are to one another, the similarity measure can impact the resulting clusters. It is thus important to consider the

distribution of one's data as well as the research question before choosing any one method. For example, there is an agglomerative method that clusters based on closest proximity; this method is better at detecting clusters when data points are distributed in a long chain of points (e.g., all in a line) than data that has points packed closely together (Milligan & Hirtle, 2012). There are many different kinds of agglomerative algorithms; however, only the most common will be described here.

The *single* linkage algorithm begins by grouping the two objects that are closest together. It then finds the next shortest distance and adds that cluster to the first cluster – or, if the next shortest distance is between two other objects, forms a new cluster containing these two. Clusters are combined based on the distance between their closest members; for this reason, this technique is sometimes called the “nearest neighbor” method (Anderberg, 1973; Hair et al., 1998). This combining process is repeated until all objects have been combined into a single cluster. The *complete* linkage method is similar to single linkage, with one notable change – rather than calculating distance based on the closest members of two clusters, it is calculated based on the farthest members (Anderberg, 1973). Despite the apparent simplicity of these methods, simulation studies have repeatedly found that the single linkage algorithm performs the worst of all the common agglomerative methods. Complete linkage typically performs slightly better than single linkage, but still tends to perform worse than other agglomerative methods (Baker, 1974; Blashfield, 1976; Milligan & Cooper, 1987; Scheibler & Schneider, 1985). One notable exception is when substantial numbers of outliers are present, in which case single linkage tends to perform the best (Milligan, 1980). Additionally, in situations in which cluster sizes are very unequal, complete linkage is typically optimal (Kuiper &

Fisher, 1975). One main advantage of the single and complete linkage methods is that they are based on rank ordering in the data matrix, and are therefore useful for ordinal data. The other agglomerative methods must be used with interval data only (Milligan & Hirtle, 2012).

Distance in the average linkage method is calculated based on the *average* distance between all objects in the first cluster to all objects in the second cluster. This distance can be used in its unweighted or weighted form (Anderberg, 1973; Hair et al., 1998; Milligan & Hirtle, 2012). Accuracy of this method tends to be mixed in simulation studies (Milligan & Cooper, 1987), with it sometimes performing the best (Kuiper & Fisher, 1975; Milligan, 1980), sometimes second best (Scheibler & Schneider, 1985), and sometimes – though rarely – worse than even complete linkage (Blashfield, 1976).

The final method that will be discussed here is the most popular and – typically – the most accurate. Ward (1963) first described a method of clustering based on within-cluster variance instead of distance. In Ward's method, group joining is based on which combinations will result in the smallest increase in within-cluster sum of squares (Anderberg, 1973; Hair et al., 1998). Simulation studies repeatedly find that Ward's algorithm provides the most accurate clustering solution, and it is thus an often-recommended procedure for cluster analysis (Blashfield, 1976; Kuiper & Fisher, 1975; Milligan, 1980; Milligan & Cooper, 1987; Scheibler & Schneider, 1985).

Divisive methods. Due to the low usage of divisive methods in the research literature, divisive algorithms will not be discussed in detail here. However, as mentioned previously, they are essentially just agglomerative algorithms in reverse (Lorr, 1983; Milligan & Cooper, 1987). For example, Edwards and Cavalli-Sforza (1965) developed a

backwards Ward's algorithm, in which clusters are split based on maintaining the smallest within-cluster variance. Although divisive methods can be more computationally complex than agglomerative algorithms, they do have the advantage of revealing the true structure of the data much sooner in the clustering process than agglomerative methods (Everitt et al., 2011).

Non-hierarchical. Whereas hierarchical methods involve a tree-like branching pattern from single observations to one large cluster (or vice versa), non-hierarchical methods – also called partitioning methods – do not. Instead, the number of clusters in which to classify observations is specified by the analyst in advance, based on theory or practicality. Thus, similarity measures take a somewhat lesser role in non-hierarchical algorithms, and the focus instead is on finding the best x -cluster solution to fit the data. To do so, a centroid (or multivariate mean – called a cluster seed in cluster analysis) is selected and all observations within a specific distance are added to the cluster associated with the cluster seed. Another cluster seed is then selected and more objects are assigned until every object is in one of the clusters. Unlike hierarchical algorithms, observations can be reassigned to different clusters throughout the clustering process (Anderberg, 1973; Hair et al., 1998; Milligan & Cooper, 1987; Milligan & Hirtle, 2012).

There are many different types of non-hierarchical clustering algorithms. Some methods select the cluster seed randomly; some use a hierarchical method as a starting point; and some require the researcher to specify the seed value. Methods also differ in how many iterations of cluster assignment they go through and the rule they use to assign objects to nearby centroids. Euclidean distance and Ward's method play a role in some non-hierarchical methods, with distance being used to assess how close a point is to a

centroid and Ward's method being used to select an initial cluster seed (Milligan & Cooper, 1987). Despite the wide variety of non-hierarchical methods available, only the most common will be discussed here. The interested reader is referred to Milligan (1980), Milligan (1996), and Milligan and Cooper (1987) for a thorough discussion and comparison of other non-hierarchical techniques.

K-means. The most common non-hierarchical technique is called *k*-means. There are several different *k*-means algorithms that have been put forth in the literature (see Milligan, 1996); however, the discussion here will focus on *k*-means methodology as described by Steinley (2003) and Tan et al. (2006). The basic technique of *k*-means involves several steps: 1. Select *k* initial centroids as cluster seeds 2. Use the squared Euclidean or city-block distance between each point and the centroids to assign each object to the nearest centroid 3. Recalculate each cluster's centroid based on the assignments 4. Reassign the points based on proximity to the new centroids 5. Continue this process until the centroids do not change anymore (Steinley 2003; Steinley, 2004; Tan et al., 2006).

The repeated iterations inherent to *k*-means is similar to the process of maximum likelihood estimation (Magidson & Vermunt, 2002), which will be described in more detail in the mixture modeling section of this paper. Also similar to maximum likelihood estimation is the possibility of reaching a locally optimal clustering solution – one that converges but is not the best, given the data – rather than a globally optimal solution. The quality of a solution is determined by the error sum of squares (SSE), which is calculated just as it would be in ANOVA or other types of known-group analyses. Whether one reaches the global optima is highly dependent upon the starting values used, so starting

values should thus be chosen very carefully and/or multiple sets of starting values should be used (Steinley, 2003; Tan et al., 2006).

Starting values can be selected using several different methods. The least common method involves the researchers selecting the centroids themselves. However, this is not typically recommended (Hartigan, 1975). One more common approach is to use random starting values. Clustering can then be accomplished either by finding a single cluster solution based on the random centroids, or by performing multiple clusterings with multiple random starting values and then selecting the solution with the smallest SSE. However, both of these methods have been shown to produce poorly optimized solutions (Milligan, 1980; Milligan & Cooper, 1987; Tan et al., 2006). A third method of selecting starting values involves using a hierarchical method, such as Ward's algorithm, to define a set number of clusters. The centroids from these clusters are then used as starting values for the k -means algorithm. This method has intuitive appeal, both because it avoids the issues caused by using random starting values and because it assists the researcher in determining how many clusters should be specified at the beginning of the analysis. Ward's method in particular has been shown to provide accurate results in past simulation studies (Milligan, 1980; Scheibler & Schneider, 1985). As a result, several theorists recommend using this technique (Milligan & Cooper, 1987; Steinley, 2003).

Comparison to hierarchical methods. There is ample evidence to suggest that k -means methods generally outperform hierarchical methods in terms of accuracy, even under extreme error conditions, *if* the starting values used are reasonable (i.e., not random). When random starting values were used, algorithm performance suffered considerably, particularly in datasets containing various levels of error perturbation

(Milligan, 1980; Milligan, 1996; Milligan & Cooper, 1987; Scheibler & Schneider, 1985). *K*-means also tends to be superior to hierarchical methods with large sample sizes, as hierarchical analyses run much less efficiently under such conditions than *k*-means does (Steinley, 2003). Additionally, hierarchical methods tend to be more influenced by outliers than *k*-means methods, which would be a distinct disadvantage in samples with a large number of outliers (Milligan, 1980). However, as already discussed, hierarchical methods have the advantage of not needing a researcher-specified number of clusters to begin the analysis, which can be a major drawback of non-hierarchical techniques. It is thus advisable to utilize hierarchical and non-hierarchical techniques together in order to benefit from the advantages of both types of methods (Hair et al., 1998).

Cluster Solution Decisions

Deciding how many clusters to ultimately retain – known as the stopping rule – is a largely subjective process. Researchers use general guidelines, theory, and practicality to guide their decision, but ultimately, there is no one “correct” answer to the question of how many clusters are inherent in the data. For this reason, it is imperative to clearly document and justify the steps one goes through in deciding on the final cluster solution (Hair et al., 1998).

Simple stopping rules. One commonly used stopping rule that can be applied to hierarchical agglomerative procedures involves an examination of a similarity value between clusters at each step. The researchers could establish a cutoff value or look for large jumps in similarity to identify a point at which the clusters that are being combined have become too dissimilar. Once that point has been determined, the researcher would then choose the number of clusters *just prior* to it in order to maximize within-cluster

similarity (Hair et al., 1998). As an example, Table 1 presents the last seven lines of an agglomeration table (the Stage and Coefficients columns) along with a researcher-generated Difference column representing the difference in magnitude from the previous stage's coefficient and the current stage's coefficient. Ordinarily, this table would extend all the way back to stage 1, with very small changes in the magnitude of the coefficients for the earlier stages. As indicated in the Table 1, there is a sizable jump in the magnitude of the coefficients from stage 90 to 91; there is an even larger jump from stage 91 to 92. It is up to the researcher to determine which magnitude jump is substantial enough to be considered the point at which the clusters have become too dissimilar. If the researcher decided the earlier (90 to 91) jump was large enough, he or she would probably posit that there are five clusters in the data. This is because the jump occurred at stage 91, and the cluster number just prior to this stage is 5 – that is, there are 5 clustering iterations between stage 90 and the end. If the researcher decided in favor of the later (91 to 92) jump, there would be four clusters for the same reason.

Another stopping rule process that applies to hierarchical agglomerative or divisive procedures is to examine a dendrogram. These graphs can be produced by many statistical software programs and illustrate the cluster combination hierarchy. Dendrograms resemble the roots of a tree, branching from a single cluster and terminating in a node that represents a single case (in the case of divisive methods) or combining with similar cases/clusters to eventually form one large cluster (in the case of agglomerative methods; Lorr, 1983; Milligan & Hirtle, 2012). In dendrograms, the height of the branches at the point of combination (or division) indicates how similar the cases or clusters being joined/divided are – the taller the branch, the less similar the clusters

joined by that branch (Milligan & Hirtle, 2012; Tan et al., 2006). Thus, the point at which the branches begin to grow abruptly taller indicates the point at which the clusters being combined are no longer very similar (Milligan & Hirtle, 2012). This information could then be used to inform the decision about the ultimate number of clusters to retain.

Complex stopping rules. Milligan and Cooper (1985) performed simulation studies examining an extensive list of statistically-based stopping rules that were independent of clustering method – that is, that could be used for either hierarchical or non-hierarchical procedures. Representing one of the most comprehensive stopping rule studies to date, (Milligan & Hirtle, 2012), Milligan and Cooper (1985) simulated data with 2, 3, 4, and 5 clusters and used each stopping rule to determine how many times the rule selected the correct number of clusters (Milligan & Cooper, 1985). Although Milligan and Cooper reviewed 30 different rules, only the most effective will be mentioned briefly here.

The most effective rule for identifying all numbers of clusters was developed by Caliński and Harabasz (1974). It utilizes the formula $[\text{trace } \mathbf{B}/(k-1)]/[\text{trace } \mathbf{W}/(n-k)]$, where n =the number of objects, k = the number of clusters in the solution, \mathbf{B} =the between SSCP matrix, and \mathbf{W} =the pooled within SSCP matrix (somewhat analogous to ANOVA). This rule correctly identified the number of clusters in a total of 390 out of 432 simulations (Milligan & Cooper, 1985).

Another stopping rule, developed by Raykowsky and Lance (1978), was extremely effective at identifying small numbers of clusters – exceeded in effectiveness only by the Caliński and Harabasz (1974) method. The formula for this rule is \bar{c}/\sqrt{k} , where \bar{c} is the average of the SSB/SST ratios for each variable on which the data were

clustered, and k is the number of clusters in the solution. The number of groups is then selected for the solution at which the value is highest – in other words, the solution that maximized between-cluster differences. In Milligan and Cooper's (1985) simulations, this formula functioned most accurately when there were only a few clusters (i.e., 2 to 3 clusters).

A few other stopping rules that bear mention are the one proposed by Mojena (1977) and Trace W , both of which are popular yet performed rather poorly in the Milligan and Cooper (1985) study. Besides Caliński and Harabasz (1974), a few other rules that consistently identified all numbers of clusters are Duda and Hart's (1973) rule, the C-Index, and Baker and Hubert's (1975) Gamma. Given the uncertain reliability of many stopping rules, it is advisable to use several of the better-performing ones when deciding on a final cluster solution (Milligan & Hirtle, 2012).

Although many of these algorithms and stopping rules are excellent tools for deciding on the final number of clusters to retain, the decision should also be informed by a theoretical framework. Do the clusters that are produced make sense from a theoretical and practical standpoint? If the researcher is intending to use the clusters in further research or analysis, will the clusters be useful? It is for this reason that collecting validity evidence for the clusters is a crucial part of cluster analysis (Hair et al., 1998; McIntyre & Blashfield, 1980; Milligan & Hirtle, 2012).

Validating Clusters

Although the clusters identified by cluster analysis are largely sample-dependent (Hair et al., 1998), there are ways to provide evidence for the possibility that they “actually” exist as opposed to just being a way to organize the sample data. There are

several highly technical validity analyses that can be applied to cluster analysis (see Tan et al., 2006); however, only the more common and easily applied will be discussed here.

Unfortunately, it is not possible to directly test whether the cluster organization mirrors the population structure, because the purpose of cluster analysis is to identify groups in a population where groups are unobserved (McIntyre & Blashfield, 1980). However, replicability of the solution would provide some validity evidence – that is, seeing whether the clusters identified in one sample appear similarly in another sample. Some researchers perform replication by “eyeballing” the similarities between two repeated cluster analyses based on different samples; however, this kind of subjectivity introduces unnecessary bias to the validation process. Instead, there are replication methods that make the validation process more empirically based (Breckenridge, 1989).

Breckenridge (1989) proposed developing a “classification rule” based on clustering assignment in one sample. The nearest centroid technique is a good classification rule to use, and has been supported in a simulation study examining its accuracy (McIntyre & Blashfield, 1989). The nearest neighbor method of cluster assignment has also been shown to be an accurate rule (Breckenridge, 1989). This rule would then be applied to a second sample, using the centroid values from the first sample. The members of cluster 1 in the first sample are then compared to the members of cluster 1 in the second sample to assess their similarity. This comparison can be facilitated with a kappa statistic, which ranges from 0 (no similarity) to 1 (complete similarity). To the extent that the parallel clusters are similar, it can be said that the cluster solution has been replicated in the second sample (Breckenridge, 1989; McIntyre & Blashfield, 1989). McIntyre and Blashfield (1980) conducted a simulation study

testing the extent to which kappa correlated with a measure of accuracy. They found a moderate to high correlation between the two measures, indicating that kappa may provide indirect support for the accuracy of a cluster solution as well as providing evidence for its stability.

Cluster solution accuracy can also be assessed by examining cluster composition based on variables that are known to differ across clusters. For example, suppose clusters in a dataset were formed using measures of help-seeking, self-acceptance, and worry. Also suppose that there is strong theoretical evidence that females tend to exhibit high scores on all three measures. If a cluster characterized by high levels of the measures contained more females than would be expected by chance (utilizing a chi-square analysis), this would provide evidence for the validity of the cluster (Hair et al., 1998).

Summary

Although cluster analysis has practical utility for social science research, it is only one of several classification analyses available. Indeed, the subjectivity and exploratory nature of cluster analysis has led many researchers to favor other, less sample-dependent analyses. Among the more popular of these alternative methods is mixture modeling.

Mixture Modeling

General Overview

Although cluster analysis was the primary classification analysis in the days before high-powered computers, mixture modeling has gained increasing popularity in recent years (Bauer & Curran, 2004; Magidson & Vermunt, 2002). The term “mixture” in the name refers to the assumption that a population may be made up of “mixtures” of unknown classes, or sub-populations, each of which can have their own probability

density functions and distributional form. In the case of continuous data, these probability density functions can be summed and appropriately weighted to create the overall population distribution (which may or may not be normally distributed). For example, each class in a skewed population could have a normal distribution; it is only the presence of multiple unobserved groups within the larger population that cause the population as a whole to be non-normal (Bauer & Curran, 2004; Pastor et al., 2007; Pastor & Gagné, 2013). The purpose of mixture modeling is to estimate distributional parameters for these latent classes. However, because there is no known categorical variable distinguishing the classes, they must be identified based on individuals' patterns of responding to the variables of interest (Bauer & Curran, 2004; Pastor & Gagné, 2013). Mixture modeling can be thought of as analogous to factor analysis, as both models are used to examine relationships among variables and to identify some underlying dimension. However, the key difference is that, whereas factor analysis is used to identify a latent *continuous* variable (factor) underlying the data, mixture modeling is used to identify a latent *categorical* variable (Pastor & Gagné, 2013).

Unlike cluster analysis, mixture modeling uses rigorous statistical measures of fit to help determine how many groups exist in a given population (Pastor et al., 2007). The researcher begins by hypothesizing about the number of classes and testing how well his or her sample data fits that model. Another model is then specified, and the fit of the data to that model is estimated and compared to the first model. This process is repeated with all specified models until the best-fitting solution is ultimately determined (Magidson & Vermunt, 2002; Pastor & Gagné, 2013). Because mixture modeling lacks some of the subjectivity of cluster analysis, it is often the preferred method of identifying underlying

classes in a sample or population, though it is not without its own limitations (Magidson & Vermunt, 2002; Pastor et al., 2007).

Initial Considerations

One important consideration for researchers is whether to approach the analysis via a direct or indirect approach (Pastor & Gagné, 2013). Researchers who adopt a direct approach assume that the classes they identify are groups that actually exist in the population. Conversely, those who adopt an indirect approach use the model as a statistical tool to accomplish something other than identifying groups they think exist in the population. One example of this indirect approach would be using mixture modeling to model a non-normal distribution that may not fit more common distributional models (Bauer, 2007). Determining one's approach ahead of time is important because violating assumptions regarding the actual existence of the identified classes may lead to erroneous conclusions later on in the analysis process (Pastor & Gagné, 2013; Lubke, 2010).

Although variable standardization was an important initial consideration when performing cluster analysis, this is not the case with mixture modeling. That is, different variable scales will not affect the classification solution like they do in cluster analysis. Given the differing views regarding the most appropriate way to standardize variables in cluster analysis, this is an advantage of mixture modeling (Magidson & Vermunt, 2002; Pastor et al., 2007).

Specifying Models

Choosing number of classes. Similar to *k*-means clustering, a requirement of mixture modeling is that the researcher specifies the number of classes in advance. However, unlike *k*-means clustering, mixture modeling allows for statistical tests of

model-data fit and comparison between models with different numbers of classes (Magidson & Vermunt, 2002). As a result, it is simple to test several models with many different numbers of classes. Often, researchers will begin their analysis with a one-class model, and continue by increasing the number of classes with each successive model. This provides the researcher with a wide variety of models from which to choose the final solution (Pastor & Gagné, 2013).

As already mentioned, a mixture model analysis will typically involve testing several models with differing numbers of hypothesized classes. Although it may seem that each model is completely separate from the others due to different numbers of specified classes, this is actually not the case. Mixture models contain a mixing proportion, which represents the proportion of the sample that is in each class – essentially weighting the solution more heavily for the larger class in determining the overall distribution. In nested models, which contain nested k and $k-1$ class solutions, the mixing proportion for the additional class has simply been set to zero for the smaller ($k-1$ class) model. Because the models are used for the same sample data with only a set of parameters separating them (one of which has just been set to zero for the $k-1$ class model) and all other parameters the same, the models are considered nested. Multiple models with the same parameterization (except the mixing proportion) can be nested within one another, allowing the researcher to compare several models with differing numbers of classes within the same analysis (Pastor & Gagné, 2013; Tofighi & Enders, 2008).

Estimating parameters. As with ANOVA and other group-based analyses, a purpose of mixture modeling is to estimate the population parameters for each class,

based on the sample data. Part of this process is the selection of the proper population and class-specific distributional form of the variables of interest. For example, it may be the case that theory suggests a negatively skewed population distribution made up of two normally distributed classes. A researcher working with this theorized population would thus specify his or her model to reflect these distributions (Pastor et al., 2007). A process called maximum likelihood (ML) estimation is often used in mixture modeling to model parameters. The purpose of ML is to identify the parameter values of the population from which the sample data were most likely obtained. Various sets of parameter values are tried out with the data, with the log likelihood (LL) representing how likely the data is under each set. The likelihood function captures the log likelihood of the data (y-axis) for various sets of parameter values (x-axis). The global maxima, or highest point of this function, captures the parameter estimates associated with the highest log likelihood. When hypothetically picturing a likelihood function having the shape of a normal curve, the global maxima would be at the peak of the curve. Unfortunately, mixture models often produce likelihood functions that have more than one peak (i.e., not as smooth as the normal curve-shaped example). Because this is the case, ML estimation may converge on a set of parameter estimates not associated with the highest log likelihood, but appears to be the highest because the estimation has gotten “stuck” on a lower peak. These lower peaks are called local maxima and are the reason that multiple estimations of the model with different random starting values are essential when performing ML estimation for mixture models (Hipp & Bauer, 2006; Pastor & Gagné, 2013; Vermunt & Magidson, 2002). Another issue that sometimes arises when attempting to converge on parameter estimates is that of singularities. A singularity occurs when a point on the

likelihood distribution spikes up to infinity, and it can cause the model to fail to converge. Sometimes beginning again with different random starting values can solve this problem, but other times it is necessary to rework the model even if it means using one that is less theoretically sound (Hipp & Bauer, 2006; Lubke, 2010; Pastor & Gagné, 2013).

When estimating mixture models, the researcher is able to constrain, or fix, various parameters (means, variances, and covariances) in the model to be equal across classes. When a parameter is constrained in this way, it is not allowed to differ across classes. In some cases, this means that the parameter must have the same value(s) for all classes. In other cases, a parameter is constrained to take on a certain value in one or more classes (e.g., a parameter is set to zero as in a latent profile model). Often, researchers will allow the means to vary across classes while constraining other parameters to be equal across classes (e.g., variances and covariances). This allows for a simpler model estimation process than a model that does not constrain any parameters (Bauer & Curran, 2004; Pastor & Gagné, 2013). However, it is important to remember that the goal is to find the *best-fitting* model, not just the one that is easiest to estimate.

Evaluating Model Fit

In order to determine how well one's sample data fits the specified model, the log likelihood (LL) or, more commonly, the -2LL is calculated. LL and -2LL are based on the extent to which the sample data are likely given the estimated parameter values of the model. LL is obtained by taking the log of the likelihood estimate, and -2LL by simply multiplying LL by -2. The closer the -2LL is to 0, the better the model fits the data (Pastor & Gagné, 2013). However, it is important to keep in mind that the -2LL is not an

absolute measure of fit – that is, it is impacted by extraneous factors such as model complexity and is thus to an extent model-dependent. Information criteria (described below) are typically used to adjust for the impact that model complexity and sample size can have on the magnitude of the LL (Henson, Reise, & Kim, 2007).

Comparing across models. Although it is useful to know how well the data fit each individual model, it is also necessary to compare the models to one another to determine relative fit. There are many ways to evaluate the relative fit of the models. The most common can be easily categorized into three groups: information criteria, likelihood ratio tests, and classification-based methods (Henson et al., 2007; Pastor & Gagné, 2013).

Information criteria (IC). Among the most popular tools for model selection are information criteria (IC) measures (Vermunt & Magidson, 2002), which are based on the log likelihood. However, they correct the LL values to adjust for more complex models and allow comparison across models (Henson et al., 2007). Commonly-used information criteria for determining model fit include the Akaike Information Criterion (AIC; Akaike, 1973), consistent AIC (CAIC; Bozdogan, 1987), Bayesian Information Criterion (BIC; Schwarz, 1978), and sample-size adjusted BIC (SSABIC; Sclove, 1987). The adjustment made to the log likelihood by these four information criteria, known as a “penalty”, is based on 1. the number of parameters that are being estimated and 2. the sample size. Generally, the AIC penalizes the LL the least, followed by the SSABIC, BIC and CAIC, although this somewhat depends on sample size (Henson et al., 2007; Tofighi & Enders, 2008). Once the chosen information criterion has been computed for all models, the model with the smallest IC is chosen as the best (Pastor & Gagné, 2013). Simulation studies have shown that the SSABIC tends to be the most accurate IC, with the AIC as

the least accurate (Henson et al., 2007; Tofighi & Enders, 2008; Yang, 2006), although one recent simulation study favored the BIC as the best, particularly with large sample sizes (i.e., $n > 500$; Nylund, Asparouhov, & Muthén, 2007). Because of this, it may be best to report several different information criteria, but rely most heavily on the SSABIC when they disagree.

Why not the chi-square difference test? In many analyses that use -2LL to assess fit (e.g., nested models in confirmatory factor analysis or logistic regression), the chi-square difference test (a.k.a. the likelihood ratio test) can be used to compare across nested models and determine which one to champion. However, this is inappropriate in mixture modeling contexts because the likelihood ratio (or the difference between the two log likelihoods) does not follow a chi-square distribution. When comparing nested mixture models, the smaller model ($k-1$) is not simply a separate model with a smaller number of classes; rather, one of the classes in the larger model (k) has been fixed at zero to produce the smaller model. As a result, the shape of the chi-square distribution for the larger model's -2LL distribution is distorted, and the difference can no longer be considered chi-square distributed. This renders the chi-square difference test an inappropriate measure of comparative fit (Lo, Mendell, & Rubin, 2001; Tofighi & Enders, 2008).

Likelihood ratio tests. Although the χ^2 difference test is inappropriate for examining k -class vs. $k-1$ class mixture models (Tofighi & Enders, 2008), there are other methods of assessing the likelihood ratio that can be used instead. One of the best known is the Lo-Mendell-Rubin test (LMR; Lo et al., 2001). Lo and colleagues corrected for the fact that the LR is not chi-square distributed by creating an adjusted distribution based on

weighted sums of chi-square values. Using the new distribution as a reference, nested models with k and $k-1$ classes can be compared based on the null hypothesis that they both fit the data equally well. A significant p -value indicates that the full (k) model fits the data better than the reduced ($k-1$) model (Tofighi & Enders, 2008). Numerous simulation studies have supported the accuracy of the LMR method in identifying well-fitting models (Henson et al., 2007; Nylund et al., 2007; Tofighi & Enders, 2008). One disadvantage of the LMR as compared to using information criteria is that the LMR can only be used to compare k -class vs. $k-1$ class nested models, while IC can compare both nested and non-nested models. Therefore, it is often best to use both LMR and IC in tandem.

Classification-based methods. Another method of assessing model fit involves determining how accurately the model classifies cases into appropriate classes, which is accomplished by calculating the posterior probability of a person's membership in each class identified by the model. These probabilities are calculated using the parameters estimated by the model and each person's actual score on the variables of interest (Pastor & Gagné, 2013; Vermunt & Magidson, 2002). In a model that does a good job classifying persons, each individual in the dataset will have a much larger posterior probability for their assigned class than for any of the other classes. Accuracy of classification can then be compared across models to determine which model classifies persons the best (Pastor & Gagné, 2013).

Classification accuracy can be used on its own to assess fit, but can also be combined with information criteria for a more robust measure (Pastor & Gagné, 2013). Two such measures – the classification likelihood information criterion (CLC) and the

integrated classification likelihood (ICL-BIC) – utilize either the -2LL or the BIC along with a classification statistic called an entropy term (E; Henson et al., 2007). E is calculated based on posterior probabilities, sample size, and number of classes. It ranges from 0 to 1, with values closer to 1 meaning that the model more accurately classifies cases than models with low E values (Henson et al., 2007; Pastor & Gagné, 2013).

Selecting the final solution. Having statistical information from which to make decisions about the appropriate number of classes for one's data is clearly a benefit of a mixture modeling approach. However, these criteria should not be the only thing on which the researcher bases final model selection (Pastor & Gagné, 2013). As with cluster analysis, the principal consideration should be whether the classes make theoretical sense. It is sometimes the case that past research suggests a particular number and configuration of classes in the population of study. In these instances, it may be best to take a more confirmatory approach to mixture modeling. This kind of approach allows the researcher to test specific hypotheses by constraining parameters in a manner consistent with theory, and may provide a more meaningful solution than would be produced by relying on statistics alone (Finch & Bronk, 2011). For example, past research may suggest the existence of three sub-populations in a larger population of college students, with Group A exhibiting much higher levels of help-seeking behavior than Group B, which in turn exhibits higher levels than Group C. The researcher can then model this constraint (Group A > Group B > Group C) to test this hypothesis.

Another consideration when choosing a model is the size and configuration of classes. Perhaps statistical criteria indicate that a 3-class solution describes the data better than a 2-class solution; however, the third class only contains a small fraction of the

sample. Not only might such a small class be more trouble to deal with than it is worth, such a situation could result in unstable parameter estimates for the small class if the sample size is not sufficiently large. As with all decisions regarding final model selection, however, theory should ultimately guide the decision of whether to retain the small class (Pastor & Gagné, 2013).

In a related vein, the researcher should also examine the patterns of variables within each class. With classification analyses, it is sometimes the case that, rather than identifying classes with qualitatively distinct patterns of responding, the analysis is simply categorizing a continuous variable. For example, a two-class solution may consist of a class with individuals who were high on all measures, and a second class with individuals who were low on all measures. While there are technically two groups of responders in this situation, such a classification would not provide any meaningful information to the researcher (McLachlan & Peel, 2000; Pastor & Gagné, 2013).

A final issue that may arise when choosing a model involves the issue of using information criteria to choose among models. As already discussed, using information criteria to choose among models involves penalizing models with more parameters – that is, more complex models. As a result, when evaluating IC, models with a large number of parameters could be rejected in favor of models with fewer parameters. Because of this, it is often advisable to present several plausible models rather than attempting to narrow the final solution down to just one model (Lubke, 2010).

Validity Evidence for Classes

Like cluster analysis, providing validity evidence for the classes identified by mixture modeling analysis is an important step in the analysis process. Replication with

different samples is always a good way to validate classification results. However, mixture modeling also provides some other, unique methods of validation that can be employed (Lubke, 2010; Pastor & Gagné, 2013).

The accuracy of a classification solution is best supported by determining if the classes relate to other variables, called correlates, in theoretically expected ways (Clark, 2010). One popular method of investigating correlate/class relationships involves assigning persons to the class for which they have the highest posterior probability, and then using the correlates and resulting groups in subsequent analyses such as ANOVA or chi-square. However, issues can arise when using the classification accuracy of the model is not strong. To illustrate, an individual who was assigned to a class because they had a posterior probability of .99 would be considered the same as an individual who was assigned to the same class with a posterior probability of .51. However, this poses obvious practical issues. This method of validation ignores the accuracy of class assignment and should thus not be used (Clark, 2010; Pastor & Gagné, 2013).

Alternatively, correlates can be included in the mixture model along with the classification variables as latent class predictors or outcomes (Clark, 2010). This approach has the disadvantage of potentially causing the classification structure to change once the correlates are included in the model (Asparouhov & Muthén, 2013; Marsh et al., 2009). Several methods have been proposed to address this issue (Asparouhov & Muthén, 2013).

One correlate-included method that also addresses the issue of class assignment accuracy is the pseudoclass drawing method (Lanza, Tan, & Bray, 2013). In this process, each case is assigned to a “pseudoclass” by randomly drawing from their posterior

probability distribution created during the mixture modeling analysis. The correlate statistics (e.g., means, variances, etc.) are then calculated after each pseudoclass draw and averaged across all pseudoclasses to get the final statistics. It is this final set of statistics that are used in analyses examining the relationship between the correlate and the classes (e.g., regression). This method has been shown to work well when classes are highly separated; however, there is an even better validity method that can be used (Asparouhov & Muthén, 2013; Pastor & Gagné, 2013; Wang, Brown, & Bandeen-Roche, 2005).

Asparouhov and Muthén (2013) described a three-step method of class validation. First, the latent classes are identified as usual. Next, a class indicator is calculated for each person, based both on the posterior probabilities as well as a term that takes assignment uncertainty into account. Finally, this modified class assignment is used in further analyses with the correlate, such as logistic regression (Asparouhov & Muthén, 2013). Lanza et al. (2013) described a similar method that used Bayesian methodology to calculate the posterior probabilities. Both methods have been shown to produce accurate results and are excellent ways of validating the classes that are identified in mixture modeling.

Comparing Mixture Modeling and Cluster Analysis

Main Differences

Clearly, there are many similarities between direct approaches to mixture modeling and cluster analysis. Their primary purpose – grouping persons based on their levels on particular variables – is identical. However, the methods by which this purpose is accomplished and the assumptions underlying the groupings are quite different (DiStefano & Kamphaus, 2006).

The major difference between cluster analysis and mixture modeling is that mixture modeling is a model-based procedure whereas cluster analysis is not. A model-based approach is based on a hypothesized model of the larger population from which the sample data is drawn (Magidson & Vermunt, 2002). In the case of mixture modeling, the theorized model is that there is a mixture of sub-populations whose distributions on the variables are characterized by a class-specific multivariate probability density function. It is the existence of these sub-populations within the larger population that are causing heterogeneity in the population (Pastor et al., 2007; Pastor & Gagné, 2012). In contrast, cluster analysis is a non-inferential procedure. This means that the identified clusters apply to the sample only; no attempt to make assumptions about groupings in the population can be made. Also, no probability density function or distribution is specified in cluster analysis as it is in any statistical model. This is also the reason that no statistical tests of the clustering solution exist for cluster analysis (Hair et al., 1998; Magidson & Vermunt, 2002; Whiteman & Loken, 2006).

Views regarding the nature and function of the class/cluster variable in each analysis are also different. In cluster analysis, groups are imposed on the data based on object similarity or proximity. The actual existence of such groups in the population is not an assumption of cluster analysis, and the clusters are not considered to result from an actual latent categorical variable. As a result, it is unsurprising that different clustering algorithms frequently result in different clustering solutions (Hair et al., 1998; Pastor, 2010; Whiteman & Loken, 2006). In contrast, in a direct approach to mixture modeling, it is assumed that there is an actual (though unobserved) categorical variable, which – depending on the parameterization employed – either moderates (in the case of freely

estimated models) or fully explains (in the case of models that impose local independence) responses on the indicator variables. Thus, rather than assigning persons to groups based on similarity to one another or proximity to the group centroid, the focus in mixture modeling (at least for researchers who opt for the direct approach) is to assign individuals to the latent group to which they most likely *actually* belong (Pastor et al., 2007; Whiteman & Loken, 2006).

Deciding Between Methods

Despite the similarity of purpose inherent in both cluster analysis and mixture modeling, their differences beg the question of which method should be used in situations where classification analysis is needed. Given the growing usage of mixture modeling techniques and the increased statistical stringency they provide (Magidson & Vermunt, 2002), there are many researchers who support the use of mixture modeling over cluster analysis (e.g., Magidson & Vermunt, 2002; Meehl, 1992; Pastor et al., 2007).

Comparative and simulation studies also often indicate that the mixture modeling provides more accurate classification than cluster analysis (DiStefano & Kamphaus, 2006; Magidson & Vermunt, 2002; Whiteman & Loken 2006). However, there are advantages and disadvantages to each method that should be considered before making a decision about which technique to use.

Cluster analysis. Although the inability to make inferences from the clustering solution to the population could be considered a disadvantage of cluster analysis, in some cases its non-inferential nature may be appropriate. Perhaps a researcher has collected questionnaire data prior to implementing an intervention in a particular classroom. The researcher would thus be interested in the sample data only, and cluster analysis may be a

flexible and useful way to group students based on their questionnaire responses. Cluster analysis' non-inferential quality may also be appropriate when a researcher is attempting to develop a theory or hypothesis about his or her data, based on the sample members. In such cases, the researcher may be more interested in the characteristics of individuals who are similar to one another than in the characteristics of an actual latent group. Thus, cluster analysis would be more suitable in this situation than would mixture modeling (Hair et al., 1998). Another advantage of cluster analysis over mixture modeling is that, because parameters are estimated using maximum likelihood estimation, mixture modeling requires large sample sizes (Enders, 2005). In situations where sample sizes are low, cluster analysis may be a better choice. Finally, unlike mixture modeling, cluster analysis does not require that a class-specific probability distribution be specified (Pastor et al., 2007). For researchers who do not have a good sense of what distribution they should choose, cluster analysis may be a better choice.

However, there are situations in which cluster analysis is at a disadvantage. The subjective nature of the decision-making process and the lack of statistical tests to assess the clustering solution are two major shortcomings of cluster analysis (DiStefano & Kamphaus, 2006). A related disadvantage is that cluster analysis will always produce clusters, even in a sample where clustering may be unnecessary or even inappropriate. This tendency has the potential to be misleading if the researcher is not aware of it, or does not collect validity evidence for the clusters (Meehl, 1992). Although this is also the case for mixture modeling – the analysis will always provide a k -class solution if one is requested – there are many more ways to tell which solution is best than there are in cluster analysis. As a final limitation, the clustering solution is completely dependent

upon the indicator variables. The addition or removal of any one variable may completely change the clustering result, which is obviously a disadvantage when attempting to draw conclusions regarding the sample of interest (Hair et al., 1998; Pastor et al., 2007).

However, this is a potential disadvantage of mixture modeling as well.

Mixture modeling. Utilizing a model-based approach like mixture modeling has the major advantage of being less subjective than cluster analysis and allowing for the application of statistical tests of model-data fit (Magidson & Vermunt, 2002). The flexibility of mixture modeling is also an important benefit, as parameters can be constrained to any degree specified by the researcher. The ability to fix parameters makes mixture modeling ideal for a more confirmatory approach to research, particularly when previous findings suggest a particular data structure (DiStefano & Kamphaus, 2006; Magidson & Vermunt, 2002; Whiteman & Loken, 2006). Another advantage of mixture modeling over cluster analysis is the lack of necessity to standardize the variables. As discussed previously, there is some disagreement regarding the best method of standardizing variables in cluster analysis (e.g., Fleiss & Zubin, 1969; Milligan & Cooper, 1988; Steinley, 2004). However, in mixture modeling, variable scaling is not an issue, and thus variables do not need to be standardized prior to running the analysis (Magidson & Vermunt, 2002). A final major advantage of mixture modeling is the possibility of fractional class membership – that is, a given individual does not absolutely belong to one class or another (Magidson & Vermunt, 2002).

One disadvantage of mixture modeling is that the number of classes must be specified in advance. The exploratory nature of cluster analysis makes it ideally suited for identifying a grouping structure when there is no previous theory to suggest one (Hair et

al., 1998; Whiteman & Loken, 2006). It is important to note that mixture modeling researchers do often run multiple models with different numbers of classes, which allows them to take an exploratory approach akin to performing *k*-means cluster analysis with different numbers of clusters (Pastor & Gagné, 2013). However, there is no mixture modeling method analogous to hierarchical cluster analysis, which can suggest the best number of clusters when the researcher has absolutely no idea where to begin. As another disadvantage of mixture modeling, one simulation study has suggested that mixture modeling may perform poorly when variable variances are unequal or when there are a large number of classes (Steinley & Brusco, 2011). A final disadvantage pertains to the necessity of specifying the class-specific distributional forms in advance of running the models. If the distributional form is misspecified, there is some danger of spurious classes being adopted – that is, the analysis may suggest a particular number of classes when in fact fewer classes actually exist (Bauer & Curran, 2004).

Given the various advantages and disadvantages inherent to both cluster analysis and mixture modeling, the current study compared the two methods. Marsh and Hau's (2007) principle of methodological synergy was utilized via an applied example of both techniques. Both substantively useful and methodologically sound, this example effectively illustrates cluster analysis and mixture modeling and, hopefully, facilitates greater understanding of the methodology involved in conducting these classification analyses.

Applied Example: Theoretical Background

For the applied example, both cluster analysis and mixture modeling were conducted using college students' scores from several different measures that relate to

student success. Identifying groups based on patterns of responding to success-related measures has utility for the higher education professional. The groupings could be used in subsequent analyses to determine the nature and extent of their relationship to student success (typically GPA), and perhaps assist in early interventions with at-risk students.

However, before students can be classified based on particular variables, the variables must be selected. The university of study, like many other universities, currently employs a variable-centered approach to predicting student success, utilizing variable-centered methods such as multiple regression. Because of this, some detail is already known regarding which variables best predict student GPA within a variable-centered analysis. Utilizing similar variables in the classification analyses is an excellent place to start, as their utility at predicting success has already been established both in practice and in previous literature. However, whereas using variable-centered analyses inherently assumes that groups are homogeneous on the variables, employing person-centered analyses allows individuals to differ across the variables. This provides information about groups and individuals that is not provided by variable-centered methods alone. An additional advantage of applying person-centered analyses to these same variables is that – as already discussed – it will be much easier to notice and examine complex interactions among the variables than by modeling interaction within a variable-centered method (e.g., regression). Having captured the complex interactions inherent in the data, the pattern profiles can then, in turn, be used in analyses such as multiple regression to predict student success.

What might the groups identified by the classification analyses look like? Perhaps some students exhibit adaptive patterns on the grouping variables – high scores on all the

“good” variables (those typically positively related to academic success) and low scores on all the “bad” ones (those typically negatively related to success). Other students may exhibit maladaptive patterns – low scores on all the “good” variables and high on all the “bad” variables. Still others may exhibit patterns that fall somewhere in the middle.

Figure 4 provides an example graph of profiles that may emerge, based on the grouping variables that will be described below. Cluster/class 1 in this example graph is the adaptive cluster – they are high on mastery approach and performance approach, and are low on performance avoidance, work avoidance and the maladaptive help-seeking orientations. In contrast, cluster/class 2 exhibits an opposite pattern, being low on the adaptive goal orientations and high on performance avoidance, work avoidance, and help-seeking. Class/cluster 3 exhibits an interesting pattern – students in this group are high on the mastery approach orientation, but are low on the performance goals and the maladaptive variables. From this example graph, it is easy to see how useful classification analyses can be, providing a quick overview of the relationships inherent in the data.

To that end, a brief theoretical background will be given for each variable in the current study before describing the analysis process. Grouping variables will be described first. These variables are used to create the clusters and classes for cluster analysis and mixture modeling. The next set of variables described are those that were used for validity evidence. The validity evidence variables are related either to student success, the grouping variables, or both. The validity variables were examined for each cluster and class, to provide evidence that the clusters and/or classes make sense from a theoretical perspective. The grouping variables that were used are achievement goal orientation

(mastery approach, performance approach, and performance avoidance), work avoidance, executive help-seeking, and help-seeking threat. The validity variables are self-acceptance, help-seeking avoidance, conscientiousness, and openness.

Grouping Variables

Goal orientation. Motivation to learn and succeed has been consistently and positively related to academic success (Elliot & McGregor, 2001; Elliot, McGregor, & Gable, 1999; Linnenbrink & Pintrich, 2002), in Anglo-American students as well as across cultures (Zusho, Pintrich, & Cortina, 2005). Robbins, Davis, Lauver, and Langley's (2004) exhaustive meta-analysis of studies examining psychosocial factors that predict college outcomes found academic motivation to be the second most powerful predictor of academic achievement, only exceeded in importance by the related concept of academic self-efficacy. Although motivation is important to success, research has also indicated that the type of motivation has a significant impact on the depth of learning and, thus, overall academic success.

Dweck (1986) described two kinds of motivational approaches: mastery and performance goal orientations. Students who endorse a mastery-approach goal orientation tend to enjoy the challenge of learning and seek to truly understand and master the material, leading to an increased likelihood that they will work hard to overcome obstacles to learning and, thus, ultimately succeed. Conversely, students who endorse a performance-approach goal orientation seek success to increase others' opinions of their ability. Consequently, students who adopt a performance orientation may tend to avoid challenges and may give up in the face of adversity. In addition to the mastery and performance distinction, an approach/avoidance component has been proposed (Elliot &

McGregor, 2001), resulting in a 2x2 framework. That is, students may approach academic situations with the goal of developing competence (*mastery-approach*), rather than concern over inability to develop competence (*mastery-avoidance*). Similarly, students may approach academic situations for the purposes of demonstrating competence (*performance-approach*) or avoiding the appearance of lack of competence (*performance-avoidance*). It is important to note that these orientations are not mutually exclusive within an individual – for example, a person may be high on both mastery approach and performance approach. Adoption of both mastery and performance approach orientations typically result in student success, though the literature is mixed (Ames, 1984; Barron & Harackiewicz, 2001; Finney, Pieper, & Barron, 2004; Petersen, Louw, & Dumont, 2008; Richardson, Bon, & Abraham, 2012). For purposes of this study, the mastery approach and performance approach orientations were considered adaptive, and the performance avoidance orientation were considered maladaptive (Barron & Harackiewicz, 2001; Elliot & McGregor, 2001). All three orientations were used as grouping variables.

Work avoidance. The concept of work avoidance pertains to a student's motivation to work hard academically. As the term implies, students who are high in work avoidance seek the path of least resistance – a way to “get by” in college without necessarily needing to learn or benefit from their experience (Brophy, 1983). Predictably, work avoidance has consistently been linked to poor academic achievement (Barron & Harackiewicz, 2003). It is also consistently negatively related to mastery goal orientations – that is, the desire to learn for learning's sake – which in turn strongly predicts academic achievement (Barron & Harackiewicz, 2003; Pieper, 2003). Previous

person-centered research has found high levels of work avoidance in profiles of students who put forth less effort in low-stakes testing contexts (Barry, Horst, Finney, & Kopp, 2010), making it an important factor to consider when predicting overall academic success.

Help-seeking behavior. Adaptive academic help-seeking behavior has been consistently related to academic success (Karabenick, 2003; Karabenick & Dembo, 2011; White & Bembenuddy, 2013). Learning to ask for help in a constructive and self-educational way is an important step in becoming a self-regulated learner. Self-regulated learners are more cognitively engaged in their learning material than non-self-regulated learners, and are thus more prone to academic success (Karabenick & Dembo, 2011; White & Bembenuddy, 2013). Adaptive help-seeking is particularly important in the college environment, where large classes are the norm and professors are less accessible than they might have been during a student's high school experience (Karabenick, 2003).

Despite the importance of engaging in help-seeking, however, some students are unwilling to seek help when they need it, whether from professors or even their peers (Karabenick & Dembo, 2011; Karabenick & Knapp, 1991). Students who do not wish to seek help may believe that asking for help is a sign of weakness or a source of embarrassment, or they may view help-seeking as a hazard to their self-esteem. These types of help-seekers (or rather, non-help-seekers) experience what Karabenick (2003) calls *help-seeking threat*. High levels of help-seeking threat are often correlated with poor academic performance. However, students who are able to overcome threatening feelings and ask for help anyway are typically more academically successful than students who do not (Karabenick & Knapp, 1991).

In contrast to not seeking help at all, some students seek help for maladaptive reasons. One such type of help-seeking is called *executive help-seeking*, and occurs when a student is asking for help in order to avoid having to expend time and effort on a problem. An executive help-seeking strategy fosters dependence on others and does not facilitate the executive help-seekers' learning and ultimate success (Karabenick, 2003; Karabenick & Knapp, 1991). It is thus unsurprising that, like help-seeking threat, students exhibiting high levels of executive help-seeking tend to perform poorly in academic settings (Karabenick, 2003).

Karabenick (2003) identified three other types of help-seeking in addition to the executive and threat types described above – instrumental help-seeking, formal help-seeking, and help-seeking avoidance. Unlike executive help-seeking, *instrumental help-seeking* is adaptive – instrumental help-seekers are seeking assistance to learn and understand the material rather than seeking someone to do the work for them. *Formal help-seeking* pertains to the source of the sought help. Individuals high on formal help-seeking look to professors or other authority figures for help whereas individuals low on formal help-seeking look to peers. Finally, *help-seeking avoidance* is similar to help-seeking threat. However, whereas help-seeking threat is merely a *reluctance* to seek help for fear of appearing ignorant or weak, those high in help-seeking avoidance *do not* seek help – whether because they are acting on feelings of help-seeking threat or some other reason. Further differentiating the two types of help-seeking, there is some indication that the two types of help-seeking (threat vs. avoidance) may differ in their relationship to sources of help (i.e., students with high levels of help-seeking threat may be more likely to seek help from informal sources whereas help-seeking avoiders may not seek help

from either formal or informal sources). However, studies do indicate a strong relationship between help-seeking threat and help-seeking avoidance (Karabenick, 2003).

Studies have also indicated there is a relationship between several types of help-seeking and the mastery approach, performance approach, and performance avoidance goal orientations. Mastery approach tends to be positively related to instrumental and formal help-seeking (Karabenick & Knapp, 1991; Roussel, Elliot, & Feltman, 2011), and negatively related to help-seeking threat, help-seeking avoidance, and executive help-seeking (Karabenick, 2003; Karabenick & Knapp, 1991). Performance approach and performance avoidance tend to be positively related to help-seeking threat, help-seeking avoidance, and executive help-seeking (Karabenick, 2003; Roussel et al., 2011).

When Karabenick (2003) used cluster analysis to investigate profiles based on all five help-seeking subscales, he found four clusters representing both strategic (high on instrumental and formal help-seeking) and non-strategic help-seeking patterns. However, a later study by Finney, Barry, Horst, and Johnston (2014) failed to replicate Karabenick's clusters, instead finding all strategic clusters that diverged only on help-seeking threat, help-seeking avoidance, and – to a lesser degree – executive help-seeking. Given that instrumental and formal help-seeking did not differentiate well among profiles, the current study investigated only the other three help-seeking variables, utilizing help-seeking threat and executive help-seeking as grouping variables and help-seeking avoidance as a validity variable. In sum, the grouping variables that were used in the current study are: mastery approach, performance approach, performance avoidance, work avoidance, executive help-seeking, and help-seeking threat.

Validity Evidence Variables

Self-acceptance. Self-acceptance – also called self-esteem, self-worth, or positive self-concept – is an individual’s feelings about his or her abilities and worth that impact one’s beliefs, decisions, or actions (Ryff, 1989). The concept of self-acceptance is usually found to have a positive impact on academic adjustment and achievement (e.g., Wang et al., 2012; Mooney, Sherman, & LoPresto, 1991). This may be because a student with a high sense of self-worth is typically motivated to maintain it, thus working harder in school and being more likely to succeed academically (Richardson et al., 2012; Robbins et al., 2004). In addition to being a motivating factor, students with a positive self-concept are more likely to believe they can succeed, thus prompting them to set attainable goals and cope effectively with any challenges they face while pursuing those goals (Chemers, Hu, & Garcia, 2001). Finally, self-acceptance can lead to general adjustment to college (Mooney et al., 1991), which in turn can have a powerful impact on eventual academic success (Strahan, 2002; Wintre et al., 2011). As a result, high levels of self-acceptance may be seen in any “adaptive” clusters/classes that may be identified in the current study.

Historically, self-acceptance has not been included in person-centered studies involving help-seeking and/or achievement goal orientation (e.g., Finney et al., 2014; Karabenick, 2003; White & Bembenutty, 2013). However, a more recent study of international students utilized self-acceptance as a grouping variable along with help-seeking and work avoidance, and found that it differentiated among clusters well (Pyburn, Horst, & Erbacher, 2014). Given its lack of widespread use, however, it was

decided to include self-acceptance as a validity variable in the current study rather than a grouping variable as in the Pyburn et al. (2014) study.

Help-seeking. As discussed above, help-seeking avoidance tends to be highly related to help-seeking threat. Additionally, both help-seeking avoidance and help-seeking threat exhibit similar relationships to goal orientation and academic success (Karabenick, 2003). Given that help-seeking avoidance tends to “hang together” (i.e., be similarly related, at least in a variable-centered sense) with help-seeking threat and executive help-seeking, help-seeking avoidance scores were used to provide supporting validity evidence for the clusters and classes found in the current study.

The Big Five. The Big Five personality factors – openness, conscientiousness, extraversion, agreeableness, and neuroticism – are well-known in psychological research, and have been investigated for their potential impact on everything from job performance (Barrick & Mount, 1991) to attachment styles (Shaver & Brennan, 1992) to vengeful tendencies (McCullough, Bellah, Kilpatrick, & Johnson, 2001). There has also been substantial research investigating their relationship to academic achievement. Results of such studies have been mixed, but fairly consistently indicate that the Big Five can have a substantial impact on academic achievement, (Trapmann, Hell, Hirn, & Schuler, 2007), in some cases even surpassing traditional academic indicators like the SAT in predicting success (Conard, 2006).

Unsurprisingly, conscientiousness is typically the factor most related to academic achievement (Poropat, 2009). Defined as the tendency to be extremely organized and success-oriented, individuals who are high in conscientiousness are naturally suited to succeed in an academic setting (Richardson, Abraham, & Bond, 2012). Studies and meta-

analyses examining the relationship between the Big Five factors and academic achievement consistently point to conscientiousness as an effective predictor of success indicators such as GPA (Conard, 2006; Poropat, 2009; Trapmann et al., 2007), so it should be considered in studies seeking to predict academic achievement.

Openness is also fairly consistently related to academic performance. Individuals who are high on this factor tend to be resourceful, forward-thinking, and insightful, characteristics that are beneficial in academic settings (Poropat, 2009; Richardson, Abraham, & Bond, 2012). Although conscientiousness is almost always the Big Five factor that is most related to academic achievement, studies often find openness to be the next strongest predictor (de Raad & Schoewenborg, 1996), although this relationship is not always significant (Trapmann et al., 2007; Richardson et al., 2012). However, overall, openness seems to be an acceptable predictor of academic success (Poropat, 2009).

Results for the other Big Five factors are inconsistent. Some studies suggest that neuroticism is negatively associated with academic achievement (de Raad & Schoewenborg, 1996) whereas others find no relationship (Huq, Rabman, & Mahmud, 1986). Similarly, extraversion may be negatively related to success (Furnham, Chamorro-Premuzic, & McDougall, 2003) or not related at all (Trapmann et al., 2007), although meta-analyses suggest that it is typically not a strong predictor (Poropat, 2009). Agreeableness is typically not related to academic achievement at all (Furnham et al., 2003; Poropat, 2009). Given these findings, the present study focused on conscientiousness and openness as validity variables for the clusters and classes, with the expectation that members of “adaptive” clusters and classes will exhibit higher levels of conscientiousness and openness than the less adaptive clusters.

Other validity variables. In addition to the variables described above (self-acceptance, help-seeking avoidance, conscientiousness, and openness), two other variables will be used as validity evidence: gender and academic major. Finding differences among clusters on known groups can provide further support for the cluster solution. For example, perhaps “adaptive” clusters may contain more females than would be expected by chance. This would provide support for the cluster, given females’ higher levels of overall academic success (DeBerard, Spielmans, & Julka, 2004) suggest that they may employ more adaptive strategies in academic success-related areas. Similarly, the clusters/classes may be split by, for example, STEM majors vs. arts/humanities, given what has been theorized about these majors’ different “cultures” (e.g., Davidson, 2008; Välimaa, 1998).

Past Research and Present Rationale

Previous studies have employed classification analyses to examine some of these variables in relationship to academic success in the past. As already discussed, Karabenick (2003) and Finney et al. (2014) both studied help-seeking from a person-centered perspective, utilizing cluster analysis and mixture modeling, respectively, to identify profiles of respondents. White and Bembenuddy (2013) also utilized cluster analysis to examine help-seeking profiles. All three of these studies employed some conceptualization of achievement goal orientation as validity evidence, as help-seeking is highly related to goal orientation (Karabenick, 2003); additionally, Finney et al. (2014) added work avoidance to the achievement goal construct when they examined validity evidence for their classes. Finally, Pyburn et al. (2014) utilized two help-seeking scales (executive and threat) and work avoidance to cluster international students; however, the

other achievement goal orientations were not included in this study and the sample was very specific (i.e., international students). To date, no studies have applied classification analyses to the achievement goal orientations and selected help-seeking scales *together* to create profiles in a non-specific college student sample. It is for this reason that the variables described above were selected for the current study.

Research Questions

Given the theoretical relationship between student success and the variables described above, as well as the aims and utility of cluster analysis and mixture modeling, the current study addressed the following research questions:

1. Are there typologies of students based on achievement goal orientation, work avoidance, and help-seeking that can be identified using both cluster analysis and mixture modeling? Are these typologies supported by validity evidence?
2. What differences will be observed in the profiles identified by cluster analysis and mixture modeling? How do the analyses' differences impact the findings?
3. Can these typologies be used to predict student success?

CHAPTER THREE

Methods

Participants and Procedure

Study participants were undergraduate college students at a mid-sized public university in the mid-Atlantic United States. All first-year undergraduate students at the university in which the current study was conducted are required to participate in an Assessment Day, which takes place a few days before the start of the semester. During Assessment Day, cognitive and non-cognitive instruments are administered to each student based on random room assignment. Assessment Day test administration is strictly standardized across rooms and testing session. All room proctors read the same instruction to students informing them about the test-taking procedures, the importance of the assessments to the university, and their right to informed consent. All proctors are trained, and each room is led by two proctors who oversee the room, distribute test materials, and answer any questions the students may have.

Assessment data from the 2009 student cohort were analyzed in the current study. The 2009 cohort was chosen because it is the most recent cohort that completed all the scales addressed in this study. Students completed the scales during first-year orientation for the fall 2009 semester; the GPA variable that served as the dependent variable for research question 3 is from the end of the fall 2009 semester – that is, it is students' GPA at the end of their first semester at the university. All students completed all the scales of interest. See Table 2 for demographic information. The gender and ethnic breakdown is typical of the university as a whole, as is the average age at time of survey completion. In order to determine whether gender and major were independent from one another in this sample, a chi-square analysis of gender by major was conducted. Results indicated more

females than expected in Education and Nursing (standardized residual $>|1.96|$; see Table 3), and more males than expected in Business/Economics and STEM majors.

Measures

Goal orientation. To address motivation, Elliot and McGregor's (2001) Achievement Goal Questionnaire (AGQ) was selected. The AGQ is an adaptation of Dweck's (1986) motivational theory of mastery versus performance achievement goals, expanded to include an approach/avoidance dichotomy within each category. The AGQ consists of four sub-scales representing the four achievement goals. Several studies have supported the four-factor structure (i.e., mastery-approach, mastery-avoidance, performance-approach, and performance-avoidance) of scores from the scale (Elliot & McGregor, 2001; Finney, Pieper & Barron, 2004). High subscale scores indicate high levels of each achievement goal orientation. For the current study, the mastery avoidance subscale was not included, both because the measurement properties of this subscale are weak and because the construct is less well-defined than the other three orientations. See Table 4 for Cronbach's alpha internal consistency reliability estimates for the current study. For sample items and more detail about the subscales used in this study, see the table in Appendix A.

Work avoidance. The work avoidance subscale utilized by Pieper (2003) and based on Harackiewicz et al. (2000) was administered. This scale contains four items pertaining to students' willingness to put forth work in their classes for the semester. One item is reverse worded. After appropriate reverse coding, high scores on the subscale indicate high levels of work avoidance. Pieper (2003) reported a Cronbach's alpha of .82

for the work avoidance scale. See Table 4 for Cronbach's alpha values for the current study.

Help-seeking. Karabenick's (2003) help-seeking scale consists of five sub-scales measuring different aspects of help-seeking. All five subscales were administered; however, only data for the executive help-seeking, help-seeking threat, and help-seeking avoidance were analyzed in the current study. High scores on the subscales indicate high levels of the help-seeking orientation. The scale's author reported Cronbach's alpha values of .78, .77, and .77 for executive help-seeking, help-seeking threat, and help-seeking avoidance, respectively. See Table 4 for Cronbach's alpha values for the current study.

Self-acceptance. The self-acceptance sub-scale of Ryff's (1989) Psychological Well-Being Scale was administered. According to the scale's author (Ryff, 1989), individuals who score highly on the self-acceptance sub-scale exhibit positive attitudes about themselves and are accepting of both their good and bad traits; low scorers tend to express unhappiness with themselves and their past. For the current study, a shortened version of the self-acceptance scale, consisting of 9 items rather than 20, was administered. Three of the items are reverse worded. After reverse scoring for these three items, high scores of the subscale indicate high levels of self-acceptance. The scale correlates moderately with other known measure of self-acceptance, and test-retest reliability for the original study was .85 (Ryff, 1989). See Table 4 for Cronbach's alpha values for the current study.

The Big Five. Although there are several measures addressing the Big Five, John, Donahue, and Kentle's (1991) Big Five Inventory (BFI) was administered in the current

study. Past research has supported the reliability and validity of this measure (e.g., John & Srivastava, 1999), and its simplicity and short length (44 items) make it ideal for administration in a university setting. Cronbach's alpha values for the BFI subscales are typically around .83 (Benet-Martínez, & John, 1998; John & Srivastava, 1999). Because the current study is focused on conscientiousness (9 items) and openness (10 items), only these subscales will be used for the current study. Four items on the conscientiousness and two items on the openness subscales are reverse worded. After reverse scoring for these items, high scores of the subscales indicate high levels of the trait. See Table 4 for Cronbach's alpha values for the current study.

Analysis

Data cleaning. There were no outliers or out-of-range responses. Not all students in the sample completed all subscales. Because there were no systematic patterns of missingness, data from 74 respondents were deleted, resulting in a final n of 1,231. See Table 4 for subscale alphas, means, standard deviations, skewness, kurtosis, and intercorrelations.

Cluster analysis. Cluster analyses were performed using IBM SPSS Version 21. Based on best practices as outlined in the literature (e.g., Milligan & Cooper, 1988; Everitt et al., 2011), subscale scores were range standardized prior to including them in the cluster analysis, and Euclidean distance measures were employed. Also as per best practices, the hierarchical agglomerative method with Ward's algorithm was utilized to identify an initial cluster solution (Milligan & Cooper, 1987), and the centroids from this solution were used as initial cluster seeds in a non-hierarchical k -means analysis (Milligan, 1980). Finally, agglomeration coefficients (Hair et al., 1998) and dendrograms

(Milligan & Hirtle, 2012) informed decisions about the number of clusters for the agglomerative method. Using the R (v.3.1.1; R Core Team, 2014) clusterSim package (Dudek, 2014), the Caliński and Harabasz (1974) stopping rule confirmed the number of clusters in a further *k*-means analysis.

To examine the validity of the cluster solution, the validity evidence variables described above served as the dependent variables in an ANOVA to determine whether certain clusters had significantly higher levels of the validity variables than other clusters. For example, because it is theorized that self-acceptance will be higher in adaptive clusters, it would be hypothesized that a cluster characterized by high levels of mastery approach and performance approach and lower levels of the other, maladaptive variables (i.e., PAV, WAV, HST, and EHS) should have significantly higher self-acceptance scores than clusters that exhibit an opposite pattern. Categorical validity variables – specifically, gender and academic major – were also included in chi-square analyses to see if there are (for example) more females in the adaptive clusters than would be expected by chance.

Mixture modeling. A series of mixture models were estimated using the same variables used for the cluster analysis. Because a multivariate normal probability distribution was used, there is a mean vector and covariance matrix for each class. There are many possible parameterizations available in mixture modeling; however, only three were selected and compared to identify the best-fitting model. In all three parameterizations, means were allowed to vary across classes. Model A freely estimated between-class variances, but constrained these variances to be equal to one another within-class, and fixed all covariances to 0. Model B freely estimated both within- and

between-class variances, and fixed all covariances to 0. Model C freely estimated within- and between-class variances, freely estimated within-class covariances, and constrained covariances to be equal across classes. One-, two-, three-, four-, and five-class models were estimated for each parameterization. Model fit was assessed via AIC, BIC, and SSABIC values; a Lo-Mendell-Rubin test; and the entropy statistic. The final model was selected by considering fit and theory. Finally, the validity variables (i.e., self-acceptance, conscientiousness, openness, and help-seeking avoidance) were included as auxiliary variables and the differences between classes were computed via the Lanza method (Asparouhov & Muthén, 2013).

CHAPTER FOUR

Results

Research Question 1a: Identifying Typologies – Cluster Analysis

Analysis. Classification variables for this study were mastery approach (MAP), performance approach (PAP), performance avoidance (PAV), work avoidance (WAV), help-seeking threat (HST), and executive help-seeking (EHS). Subscale scores were range standardized prior to analysis using one of the equations suggested by Milligan and Cooper (1988), namely:

$$\frac{x - \text{Min}(x)}{\text{Max}(x) - \text{Min}(x)}$$

After range standardization, hierarchical cluster analysis was performed utilizing squared Euclidean distance and Ward's algorithm; the last ten lines of the agglomeration coefficient table can be seen in Table 5. Both the dendrogram and agglomeration coefficients suggested a three-cluster solution, which can be seen in Figure 5. Note that subscale *z-scores* are graphed in this figure instead of *raw scores* or *range standardized scores*. Not only does graphing *z-scores* eliminate the potential confusion of different response scales, but it also allows the subscale mean of each cluster to be compared to the other subscale means more easily. However, it does make it important to remember that these comparisons are relative and do not portray the magnitude of the means. Because the analysis suggested three clusters, the three-cluster solution's cluster assignment variable was saved, along with a two- and four-cluster solution for further testing in the *k*-means analysis.

The centroids from the hierarchical solutions were used as initial cluster seeds in two-, three-, and four-cluster *k*-means analyses. Caliński and Harabasz's (1974) pseudo-*F*

statistic was 2.39 for the two-cluster solution, 25.37 for the three-cluster solution, and 19.58 for the four-cluster solution, indicating that the three-cluster solution was the best (as it had the largest pseudo- F value). This was supported by the agglomerative analysis findings. This final solution is presented graphically in Figure 6. As with Figure 5, note that cluster means are presented as z-scores. Also note from Figures 5 and 6 that the three-cluster agglomerative and k -means solutions are very similar, which further supports the choice of three clusters for the final k -means solution.

Description of clusters. As can be seen in Figure 6, the three clusters exhibited distinct patterns of means (see Table 6 for raw means by cluster). Students in *Cluster 1*, which was the middle-sized cluster with 420 students, were high on the goal orientation variables (MAP, PAP, and PAV) relative to the other clusters, and were low (though not always the lowest) on WAV, HST, and EHS variables. *Cluster 2* was the smallest cluster with 340 students. Despite being the second highest scorers on MAP, students in this cluster were still slightly below the overall sample mean on MAP. Cluster 2 was the lowest on PAP, PAV, and HST, and was just above Cluster 1 on WAV. Finally, *Cluster 3* – the largest at 471 members – was lowest on MAP, slightly below the overall mean on PAP, and the highest of all the clusters on WAV, HST, and EHS. However, they were at the overall mean on PAV, and still lower than Cluster 1.

Research Question 1b: Validity Evidence – Cluster Analysis

Continuous validity variables. The second part of research question 1 addressed whether the cluster solution was supported by validity evidence. The continuous validity variables – help-seeking avoidance (HSA), conscientiousness, openness, and self-acceptance – served as the dependent variables in ANOVAs with the cluster

identification variable as the grouping variable (see Table 7). There were no significant differences between Clusters 1 and 2 for any of the validity variables. However, when compared to Clusters 1 and 2, Cluster 3 reported significantly higher levels of help-seeking avoidance ($\eta^2 = .19$) and significantly lower levels of the other three variables. These findings supported the distinctiveness of Cluster 3.

Categorical validity variables. Chi-square analyses were conducted to examine the distribution of gender and major across clusters. Cells with standardized residuals greater than 1.96 were considered statistically significant. Cluster 3 consisted of more males than would be expected by chance, whereas Clusters 1 and 2 consisted of more females than would be expected by chance ($\chi^2(2) = 36.92, p < .001$).

The chi-square by major was also statistically significant, $\chi^2(14) = 47.35, p < .001$. Results are presented in Table 8. There were more Nursing majors than expected in Cluster 1 and fewer than expected in Cluster 2. Cluster 2 consisted of more Social Sciences and Education majors than expected. Finally, there were more Business/Economics students in Cluster 3 than would be expected by chance. Overall, given the distribution of observed vs. expected values among the three clusters, it appears that Cluster 1 consisted mainly of “hard science” majors (Nursing). Cluster 2 consisted mainly of Social Sciences and Education; and Cluster 3 consisted mainly of Business/Economics majors.

In summary, the continuous validity variables strongly supported a distinct Cluster 3. They also supported – though less convincingly – a distinction between Clusters 1 and 2. This distinction was borne out more clearly in the chi-square results by major than other external validity criteria.

Research Question 1a: Identifying Typologies – Mixture Modeling

Analysis. This research question pertained to the identification of profiles based on the classification variables (MAP, PAP, PAV, WAV, HST, and EHS) using mixture modeling. One-, two-, three-, four-, and five-class models were estimated for each of the three mixture modeling parameterizations in order to explore a wide range of possibilities while also maintaining a manageable number of classes. Fit indices for the models can be seen in Table 9. The three-, four-, and five-class solutions for Model B (freely estimated within- and between-class variances, covariances set to 0) did not appear stable, given that the log-likelihood did not replicate. The same was true for the four- and five-class Model C solution (freely estimated within- and between-class variances, freely estimated within-class covariances, and constrained between-class covariances). None of the unstable models were interpreted.

Of the interpreted models, the three-class Model C had the lowest ICs of all the solutions. It also had relatively good entropy in comparison to the other models, and the LMR test indicated that the *three*-class Model C was a better fit than the *two*-class Model C. Thus, the three-class Model C was championed (Henson et al., 2007; Tofighi & Enders, 2008).

Description of classes. Class means (raw metrics) on the classification variables are presented in Table 10, and variance/covariance matrices in Table 11; standardized means are graphed in Figure 7 (note that Figure 7 means are based on modal assignment; graphed means are thus approximate). *Class 1* (the middle-sized class with 239 students) was high on MAP, PAP and PAV, just below the overall sample mean on WAV, and at the overall mean on HST and EHS. *Class 2*, the smallest at 184 students, was in the

middle of the three classes on MAP, and had means on WAV and HST that were virtually identical to Class 1. Class 2 was also lowest on PAP, PAV, and EHS. Finally, *Class 3* – by far the largest at 808 students – was characterized by the lowest levels of MAP, levels just above Class 2 on PAP, and levels of PAV, WAV, HST, and EHS that were around the overall sample mean.

Distinguishing Classes 1 and 2 were their scores on the Performance variables (PAP and PAV). This distinction was much clearer in the mixture modeling solution than it was in the cluster analysis solution. Class 1 was very clearly high on the Performance variables in addition to MAP; in contrast, Class 2 was almost as high on MAP but was the lowest on the Performance variables of any of the three classes. Classes 1 and 2 also diverged on EHS. Class 3's profile was clearly different from Classes 1 and 2.

Research Question 1b: Validity Evidence – Mixture Modeling

Continuous validity variables. This research question addressed whether the mixture modeling classes were supported by validity evidence. In order to provide this validity evidence, the validity variables examined for the cluster analysis solution (help-seeking avoidance, conscientiousness, openness, and self-acceptance) were entered in the mixture modeling analysis as auxiliary variables. Chi-square comparisons of validity variables means across classes are presented in Table 10; note that these class means (for all the variables) were computed using information from posterior probabilities (i.e., the Lanza method; Asparouhov & Muthén, 2013) rather than modal assignment. Classes 1 and 2 statistically significantly differed from each other on openness (with the Class 1 mean being lower), but not on any of the other auxiliary variables. This lack of difference on the other variables suggests that the distinction between Classes 1 and 2 may be

weaker than for Classes 1 and 2 vs. 3, at least given the validity variables examined here. In contrast, Class 3 was characterized by significantly higher levels of help-seeking avoidance than the other two classes, and also had significantly lower levels of conscientiousness, openness, and self-acceptance than the other classes.

Categorical validity variables. Categorical validity variables were also entered as auxiliary variables in the mixture modelling analysis; thus, the chi square analysis results reported here were computed using the Lanza method with posterior probabilities rather than modal assignment. The overall chi-square analysis of class by gender was significant, $\chi^2(2) = 7.03, p = .030$, as was the Class 1 vs. Class 3 comparison, $\chi^2(1) = 6.88, p = .009$. Class 2 did not significantly differ from the other classes in terms of gender distribution. Because the auxiliary output does not provide observed vs. expected information for the groups, predicted probabilities were examined instead, and compared to chance probabilities (taking the proportion of males vs. females in the sample into account). The probability of being female was higher than expected by chance in Class 1 and lower than expected by chance in Class 3; conversely, the probability of being male was lower than expected by chance in Class 1 and higher than expected by chance in Class 3.

Employing the Lanza method, a chi-square by major was also significant, $\chi^2 = 102.75, p < .001$, and all classes significantly differed from one another in terms of major distribution (Class 1 vs. 2 $\chi^2(7) = 30.46, p < .001$; Class 1 vs. 3 $\chi^2(7) = 29.51, p < .001$; Class 2 vs. 3 $\chi^2(7) = 50.81, p < .001$). Looking at predicted probabilities vs. chance probabilities, there were proportionately more STEM and Nursing majors in Class 1,

proportionately more Arts and Humanities majors is Class 2, and proportionately more Undeclared majors in Class 3.

Research Question 2: Differences between Profiles

This research question addressed the differences observed between the cluster analysis and mixture modeling profiles. A classification table of modally assigned class-to-cluster assignment can be seen in Table 12. In terms of majority assignment, the clusters and classes tended to match – that is, the majority (81%) of students in Class 1 were assigned to Cluster 1, the majority (60%) of students in Class 2 were assigned to Cluster 2, and the majority (49%) of students in Class 3 were assigned to Cluster 3. Additionally, a chi-square analysis of (non-modal) class assignment by cluster assignment, computed via the Lanza method in Mplus (Asparouhov & Muthén, 2014), was significant ($\chi^2(4) = 48,517,672.0, p < .001$). Note the magnitude of the chi-square value. Given that the chi-square null is approximately equal to the degrees of freedom, the chi-square obtained here was relatively enormous. This casts some doubt on the findings, particularly given the fact that there was some difficulty interpreting the Mplus auxiliary output. Some output values were listed as “*****”, which the software developers indicated meant the value was too large to print. Thus, this large chi-square should be interpreted cautiously. Additionally, despite the significant chi-square analysis, there were still areas of considerable non-overlap in cluster-to-class assignment, particularly for Cluster/Class 3. That is, Class 3 ($n = 808$ – the largest class) contained 183 students who were assigned to Cluster 1 and 229 students assigned to Cluster 2.

Despite this lack of overlap, however, the general pattern of mixture modeling class profiles was still similar to the cluster analysis profiles (compare Figures 6 and 7).

Like Cluster 1, Class 1 was high on MAP, PAP, and PAV and low on WAV, HST, and EHS, relative to the other classes. Although less distinct for the mixture modeling solution versus the cluster solution, the overall ranking of the classes on MAP, PAP, PAV, and EHS was also the same for the clusters and classes (compare Tables 6 and 10). However, despite the similarity of ranking, the overall distinction among the three classes was much less defined for HST, EHS, and (to a lesser extent) WAV than it was among the clusters. Specifically, there were virtually no differences among the classes on HST, whereas Cluster 3 was much higher on HST than the other clusters in the cluster analysis. Thus, unlike the clustering solution, the WAV, HST, and EHS variables could not be used to discriminate across classes. Rather, the class profiles were more differentiated by the goal orientation variables (MAP, PAP, and PAV) than the cluster profiles were.

Additionally, consideration of the goal orientation means indicates that Classes 1 and 2 were more clearly qualitatively distinct than Clusters 1 and 2. Classes 1 and 2 were similar on MAP, but Class 1 was a *high performance* group – high on both PAP and PAV – whereas Class 2 was a *low performance* group. Although the cluster analysis solution showed a similar pattern – Cluster 1 was high on PAP and PAV and Cluster 2 was low on both variables – the clusters' MAP scores were much more disparate, which makes the high performance/low performance dichotomy less striking. Overall, the corresponding classes and clusters exhibited similar patterns of means, but with differences in terms of cluster-to-class assignment, relative magnitude of means, and distinction among profiles.

Research Question 3: Predicting GPAs with Profiles

Research question 3 addressed whether the profiles from the cluster analysis and mixture modeling could be used to predict students' GPA. As mentioned previously, the

subscale scores that were used to identify the clusters and classes were collected from entering first-year students at the beginning of the fall 2009 semester, prior to the beginning of classes at the university. The GPA data were from the end of the fall 2009 semester; therefore, the regression analyses tested whether the clusters and classes predicted end-of-first-semester GPA. As GPA data were not available for 14 of the 1,231 students, data from these 14 students were not included in the analysis.

Because research question 3 was concerned with using student profiles to predict GPA, the dummy-coded class and cluster variables were entered into a multiple regression analysis. Non-nested models were estimated and compared first. Then, in order to see if the mixture modeling solution provided any additional information above and beyond the cluster analysis solution, the variables were entered hierarchically – dummy-coded clusters first, followed by the dummy-coded classes. Analyses were also conducted entering class first, followed by cluster. Because the order of the steps was simply switched, only the step 1 and 2 R^2 values (step 1 $R^2 = .013, p < .001$; step 2 $R^2_{\text{change}} = .003, p < .001$) were different from what has is described below (compare to Table 13). However, because the clusters did not explain a significance amount of variance above and beyond what was explained by the classes, it was more informative to enter cluster first.

It should be noted that the class identification variable was based on modal assignment – that is, each person was assigned to class for which they had the highest posterior probability. As already discussed, there are issues with this method of class assignment; however, it was the best and simplest method available if the classes were to be used as variables, as they were for this regression analysis.

Non-nested regression models. The regression analysis was first conducted via a comparison of two non-nested models' predictive ability – one using cluster membership to predict GPA and the other using class membership to predict GPA. The cluster model explained a statistically, but not practically, significant amount of variance in GPA ($R^2 = .005$, $p = .045$); the class model explained more variance in GPA ($R^2 = .013$, $p < .001$), but was also not practically significant. Although the class model explained more variance in GPA than the cluster model, Steiger's test of dependent correlations (Steiger, 1980) indicated that there were no significant differences between the two models ($z = -1.18$). That is, the cluster model did not predict GPA significantly better than the class model, and vice versa.

Nested regression models. For the nested regression model, there was some initial concern over the possible issue of multicollinearity between the class and cluster variables. However, as the *phi* coefficient between the two variables was .57, the correlation was not deemed large enough to warrant multicollinearity concerns (Tabachnick & Fidell, 2013). Additionally, tolerance values for the dummy coded class and cluster variables were all above .40, and many were higher than .60. This indicated that there was not an undue amount of collinearity among the variables. It should be noted, however, that the class and cluster variables were more highly correlated with each other than they were with GPA (cluster with GPA $r = -.065$; class with GPA $r = -.046$).

Cluster/Class 3 as comparison group. For the first nested regression analysis, Cluster and Class 3 served as the comparison group (i.e., the group coded 0). Thus, the dummy coded variables representing Clusters 1 and 2 were entered first, followed by the dummy coded variables representing Classes 1 and 2, and finally the interaction terms

(i.e., two-, three-, and four-way). Results can be seen in Table 13. Please note that the sr^2 values in step 2 provide the same information as they would have had the class variable been entered first.

As can be seen from Table 13, the interaction step was not significant ($R^2_{\text{change}} = .001$, $F_{\text{change}} = .374$, $p = .772$), suggesting that the interaction terms did not explain a significant amount of variance above and beyond the cluster and class variables. Step 1 – which entered the cluster variables – explained a significant amount of variance in GPA, $R^2 = .005$, $F = 3.112$, $p = .045$. The significant b -values indicate that both clusters' means were significantly higher than Cluster 3 (because the b 's are positive). However, the increment of variance explained by step 2 (which entered the class variables) above and beyond the variables entered in step 1 was also significant ($R^2_{\text{change}} = .011$, $F_{\text{change}} = .6549$, $p = .001$), meaning that the classes explained a significant amount of variance in GPA above and beyond what was explained by the clusters. The b -values and sr^2 's for step 2 indicate that the increment increase was carried entirely by the difference between Class 2 and 3 GPA; given the other predictors in the model, Class 2's mean GPA was statistically significantly higher than Class 3's. Class 1's b indicated that there was no difference between the Class 1 mean and the Class 3 mean, controlling for the cluster variables. Thus, in summary, although the clusters explained a significant amount of variance in GPA, Class 2 explained even more, above and beyond what was explained by the clusters.

Despite this statistical significance, the effect sizes remained relatively small. Class 2's sr^2 was only .01, indicating that it explained 1% of the variance in GPA above and beyond the other predictors in the model. The overall variance explained by the

model with both clusters and classes was also small at 1.6%, and the increase in explanatory power from step 1 (clusters) to step 2 (classes) was only 1.1%. Thus, although the classes (and more specifically, Class 2) were able to explain a significant amount of variance in GPA above and beyond what was explained by the clusters, in effect size terms this explanatory power was relatively weak.

Cluster/Class 2 as comparison group. A regression analysis was also conducted with Cluster/Class 2 serving as the comparison group (i.e., group coded 0) rather than Cluster/Class 3. Results are presented in Table 14. Because the variables were entered in the same order (clusters first, then classes, then interaction terms) the statistics for each step (i.e., R^2 , F -values, etc.) were the same. However, the b -values and sr^2 's were of particular interest. Notably, Cluster 1's b -value indicated that the Cluster 1 GPA was not significantly different from Cluster 2's GPA. In contrast, Class 1's b -value indicated that the Class 1 GPA was significantly different (specifically, lower) than Class 2's GPA. However, as with the previous regression analysis, the effect sizes were extremely small. The Class 1 sr^2 indicated that Class 1 explained only .7% of the variance in GPA above and beyond what was explained by the other predictors, and Class 2 explained only 1%.

Cohen's d comparisons. Because using modal assignment to assign individuals to mixture modeling classes is not considered best practice, GPA analysis was also conducted using GPA as an auxiliary variable (Lanza method; Asparouhov & Muthén, 2014). Entering GPA into the mixture model analysis in this way eliminates the need to assign individuals to one class or the other (i.e., fractional class membership is maintained), thus avoiding the difficulties that can arise from using modal assignment to assign individuals to classes. Using GPA as an auxiliary variable also allows for a class-

to-class comparison of means that provides the same information provided by entering categorical predictors into a regression analysis.

Class-to-class comparison chi-square values can be seen in Table 15. Results were the same as the regression analysis. Class 2's GPA was significantly higher than Class 1 and 3's GPA. There were no significant differences between Class 1 and Class 3.

However, unlike the regression analysis, effect size differences were larger than they were when modal assignment was used. Cohen's d differences for the modally assigned classes and the fractional membership classes are presented in Table 15. Whereas the Class 1 vs. 2 and Class 2 vs. 3 comparisons resulted in small effect sizes for the modally assigned classes (according to Cohen's benchmarks; Cohen, 1992), the effect sizes were medium for the fractional membership classes. Table 15 also presents the d values for the cluster-to-cluster mean comparisons. As with the regression analyses, the effect sizes are extremely small, indicating that the clusters did not significantly differ on GPA. In addition to d , r^2 values were also calculated for the fractional membership class comparisons in order to discuss them in variance explained terms and compare them to the clusters' and modal classes' r^2 values. The r^2 for Class 1 vs. 2 was .10 and for Class 2 vs. 3 was .09 (both large effects); for Class 1 vs. 3 r^2 was .00. Thus, the classes explained significantly more variance in GPA than the clusters, as was found in the regression analyses. However, the fractional membership classes explain even more variance in GPA (in terms of effect size comparison) than the modally assigned classes.

CHAPTER FIVE

Discussion

Brief Overview

Research questions. This study was designed to address three research questions. The first question asked whether there were typologies of students based on achievement goal orientation, work avoidance, and help-seeking that could be identified using both cluster analysis and mixture modeling. This question further explored the validity of these potential profiles, based upon differences on several continuous and categorical validity variables. The second research question pertained to the differences that would be observed in the cluster analysis and mixture modeling profiles, and addressed how these differences would impact the final solutions. Finally, the third research question involved using the profiles to predict student success.

Variables of interest. Students were classified on six variables: mastery approach (MAP), performance approach (PAP), performance avoidance (PAV), work avoidance (WAV), help-seeking threat (HST), and executive help-seeking (EHS). Generally, MAP and PAP tend to be adaptive orientations, in terms of motivation and academic success; whereas the PAV orientation tends to relate to less adaptive, or self-regulated, learning strategies (Barron & Harackiewicz, 2001; Elliot & McGregor, 2001). Thus, one would expect a profile characterized by high scores on MAP and PAP but low scores on PAV to be academically successful. Additionally, work avoidance (Barron & Harackiewicz, 2003) and the two help-seeking scales (Karabenick, 2003) tend to be negatively related to student success, suggesting that academically successful students would be more likely to exhibit low levels of these variables.

In addition to the classification variables, this study examined several other variables to provide validity evidence for the clusters and classes – help-seeking avoidance (HSA), self-acceptance, and the Big Five traits of conscientiousness and openness. Help-seeking avoidance has been negatively related to academic success (Karabenick, 2003) whereas self-acceptance (Strahan, 2002; Wintre et al., 2011), conscientiousness (Poropat, 2009), and openness (de Raad & Schoewenburg, 1996) are typically positively related to academic success. Thus, it would be expected to see low levels of HSA and high levels of self-acceptance, conscientiousness, and openness in clusters/classes that display adaptive patterns of means on the classification variables.

Qualitative Distinction of Profiles: Cluster Analysis

Interpretation of clusters. Figure 6 provides a visual comparison of the three clusters identified by the cluster analysis. Given what past research has suggested about which classification variables are most related to adaptive learning strategies, it would seem that *Cluster 1* exhibited the most adaptive profile. Students in this cluster were high on MAP and PAP and relatively low – though not always the lowest – on WAV, HST, and EHS. However, this cluster was also high on the less adaptive PAV variable. Thus, Cluster 1 was characterized by high goal orientation scores and low WAV and help-seeking scores. *Cluster 2*'s pattern is difficult to characterize. Despite being slightly below the mean on MAP, this cluster still scored higher on MAP than Cluster 3. However, they were also the cluster that scored lowest on PAP (an adaptive variable) and PAV, HST, and EHS (less adaptive variables). Thus, Cluster 2 exhibited more adaptive characteristics (low on PAV, WAV, HST, and EHS) than maladaptive ones (low on MAP and PAP) when compared to Cluster 1. Finally, *Cluster 3* exhibited a pattern somewhat

opposite to Cluster 1. Cluster 3 was the lowest on MAP, near the mean on PAP and PAV, and was relatively high on WAV, HST, and EHS. Characterized by low MAP scores and high scores on the last three maladaptive variables, this cluster could be characterized as having the least adaptive profile.

Validity evidence. The validity evidence supported some of these characterizations of the clusters, in terms of adaptive learning strategies. Cluster 3 means on the validity variables (HSA and self-acceptance, conscientiousness, and openness) were significantly different from Cluster 1 and 2 means. As can be seen in Tables 6 and 7, significant differences were in the expected direction – students in Cluster 3 scored significantly higher on the less adaptive variable (help-seeking avoidance) and lower on the adaptive variables (conscientiousness, openness, and self-acceptance) than students in other clusters. Given that Cluster 3 exhibited the least adaptive pattern of means on the classification variables, these differences make theoretical sense.

However, there were no significant differences between Clusters 1 and 2 on the continuous validity variable means (HSA, self-acceptance, conscientiousness, and openness). This lack of difference was puzzling, particularly given the relatively wide disparity between these clusters on the PAP and PAV variables. Moreover, as noted in Table 6, mean validity variable scores between the two clusters were virtually identical. This is unsurprising for help-seeking avoidance; Clusters 1 and 2's scores on the other two help-seeking scales (help-seeking threat and executive help-seeking) were extremely similar, and help-seeking research has found that help-seeking avoidance tends to “hang together” with help-seeking threat (Karabenick, 2003). But what about the other validity variables?

One explanation for why Clusters 1 and 2 are not dissimilar on the other validity variables (conscientiousness, openness, and self-acceptance) is that these variables may be more related to the clustering variables on which Clusters 1 and 2 are *similar* (i.e., WAV, HST, and EHS) than they are to the variables on which they are *different* (i.e., MAP, PAP, and PAV). If this were the case, it would make sense for Clusters 1 and 2 to be similar on the external validity criteria because they are also similar on WAV, HST, and EHS. The correlations in Table 4 partially support this idea. Correlations between the three validity variables and the PAP and PAV variables are low; except for the correlation between Conscientiousness and PAP, they are smaller than $\pm .1$. In contrast, correlations between the validity variables and WAV, HST, and EHS are higher (with the exception of the correlation between openness and HST, they are all around .2 or above). However, conscientiousness, openness, and self-acceptance are also moderately correlated with MAP (the second-highest correlations after their correlation with EHS). Examination of the MAP means for Clusters 1 and 2 reveal that the clusters are less dissimilar on MAP than they are on PAP and PAV, which may explain why the strong correlation between MAP and the validity variables did not result in significant differences between Clusters 1 and 2.

The categorical validity variables also spoke to the qualitative distinctions among the clusters. There were more females than expected in Clusters 1 and 2 (the clusters with the more adaptive mean patterns) and fewer than expected in the less adaptive Cluster 3. More interesting, however, was the major distribution across clusters. *Cluster 1* – which had the most adaptive configuration – consisted of more Nursing majors than expected by chance. The prevalence of “hard” science majors in Cluster 1 is unsurprising. Students in

majors like Nursing typically experience more exacting academic standards than students in other majors, perhaps necessitating more adaptive academic strategies. This also explains the high performance scores (PAP and PAV), as students in these majors may be seeking to perform well as per external criteria (e.g., nursing board examinations) as much as they are seeking to master their course material. *Cluster 2* included more Social Sciences and Education majors than expected. With less exacting academic standards than the “hard” sciences, the *low* performance scores seen in this cluster make more sense. Still in the middle on mastery (relative to the other clusters) and low on WAV, HST, and EHS, the pattern seen in Cluster 2 may in fact be adaptive for Social Science and Education majors, who do not need to worry as much about external standards.

Cluster 3 – the cluster with the least adaptive configuration – consisted of more Business/Economics majors than would be expected by chance. The explanation for this is less forthcoming than it was for the other clusters. Business/Economics is arguably different in terms of academic culture than the sciences and education; perhaps the academic strategies that are valued in the Business world are different from those valued in other fields. Alternatively, there are more males in Cluster 3, and there are also more males in Business/Economics majors than expected by chance (see Table 3). Thus, it may be gender that is driving more Business/Economics majors to be assigned to Cluster 3, or it could be major that results in more males being assigned to Cluster 3. Moreover, it is important to keep in mind that students completed these measures before they had actually completed any coursework; thus, the question becomes whether they exhibited these profiles because of their chosen major, or whether they chose their major because they exhibited these profiles. Despite the uncertainty regarding an interpretation of this

result, the clear major-specific distinctions among the clusters provided validity evidence for the championed three cluster solution.

Conclusions. In summary, the evidence supported a distinct Cluster 3 (the least adaptive profile). Despite the disparity in goal orientation variable means for Clusters 1 and 2, the continuous validity variables did not distinguish well between the two clusters, although correlations between the validity variables and the variables on which Clusters 1 and 2 were most similar (WAV, HST, and EHS) may explain this lack of difference. However, the categorical validity evidence provided stronger support for a distinct Cluster 2. Although both Clusters 1 and 2 consisted of more females and fewer males than expected by chance, the clusters were more clearly distinguished by distribution of majors – “hard” sciences in Cluster 1, “soft” sciences in Cluster 2, and Business/Economics majors in Cluster 3.

Future research should further investigate the relationship between Business/Economics majors and the patterns observed in Cluster 3. Why did the cluster with the least adaptive profile include more Business students than expected? Research into academic strategies espoused by Business majors would be an excellent place to start. Overall, major provided clear distinctions among the clusters observed in this study, but further research may provide more insight. The findings also suggest that additional research on how the goal orientation variables relate to self-acceptance, conscientiousness, and openness is warranted. Moreover, prior to making strong claims about the “existence” of clusters, replication studies are recommended.

Qualitative Distinction of Profiles: Mixture Modeling

In addition to cluster analysis, a series of mixture models were estimated using the classification variables (MAP, PAP, PAV, WAV, HST, and EHS).

Interpretation of classes. See Figure 7 for a visual comparison of the three mixture modeling classes. Unlike the clustering solution, there was no class that exhibited a clearly adaptive pattern of means on the variables. *Class 1* was high on the adaptive MAP and PAP variables and was below the mean on WAV. However, this class was also highest on PAV and at the mean on HST and EHS. Class 1 was technically the lowest on WAV and HST, but as can be seen in Figure 7 and Table 10, the difference between all of the classes on HST and between Class 1 and 2 on WAV was virtually nil. *Class 2* was also high on MAP – though slightly below Class 1 – was low on PAV and EHS, and was just below the mean on WAV. However, Class 2 was also low on PAP and at the mean on HST. It can thus be said that Classes 1 and 2 in some ways both exhibited patterns of means that were adaptive, with neither one exhibiting a completely adaptive pattern. Class 1 was high on both MAP and PAP but was also relatively high on the less adaptive variables, PAV and EHS; Class 2, in contrast, was high on MAP but not PAP, but was also lower on PAV and EHS than Class 1. *Class 3* exhibited the least adaptive pattern of means, with the lowest mean on MAP and the highest means on WAV and EHS. The Class 3 mean was higher than Class 2 on PAP and PAV, but was still below the overall sample mean. An important note when considering all the classes together is the utter lack of differences on HST. All three classes were at the mean on this variable, indicating that it did not aid in distinguishing among the three classes.

Validity evidence. Of the continuous validity variables (HSA, self-acceptance, conscientiousness, and openness), Classes 1 and 2 only differed from one another on the personality trait of openness, with Class 1 scoring lower than Class 2. However, Class 3 – the class with the least adaptive pattern of means – significantly differed from Classes 1 and 2 on all the external criteria. These differences were in the expected direction, as Class 3 had a higher HSA mean (less adaptive variable) and lower self-acceptance, conscientiousness, and openness scores (more adaptive variables) than the other classes (see Table 10). One possible explanation for the lack of differences between Classes 1 and 2 on HSA, self-acceptance, and conscientiousness is similar to the explanation provided for the lack of differences on these variables for the clustering solution. Like the clusters, Classes 1 and 2 are similar on WAV, HST, and EHS. If the three validity variables were more related to WAV, HST, and EHS than they were to the other variables – which is partially supported by the correlation table – it would make sense that Classes 1 and 2 were not differentiated on the validity variables. However, this explanation is not as convincing as it was for the clustering solution, given that Classes 1 and 2 diverge more obviously on EHS than Clusters 1 and 2 did.

As with the clusters, there were more females than expected by chance in Class 1, which exhibited a moderately adaptive academic pattern (DeBerard, Spielmans, & Julka, 2004). However, Class 2 did not consist of more females than expected by chance, even though Class 2's pattern was similarly adaptive to Class 1's. The reason for this may lie in the chi-square results by major. Similar to Cluster 1, Class 1 was represented by more Nursing majors than expected by chance – an academic population that is typically overwhelmingly female. Indeed, as noted in Table 3, there were significantly more

female Nursing majors than would be expected by chance. Furthermore, unlike *Cluster 2* (which included more Education majors than expected by chance), *Class 2* was characterized by more Arts and Humanities majors than expected by chance. Although there were significantly more female Education majors than expected by chance (explaining the significantly higher number of females in *Cluster 2*), there were *not* significantly more of either gender in the Arts and Humanities (explaining the lack of gender differences in *Class 2*). These results are telling, and suggest that it may be the case that the gender distribution for both the clusters and the classes may be a function of the major distribution. However, this does not explain the major distribution in *Class 3*, in which there were more Undeclared majors and males than expected by chance. Table 3 indicates that there were *not* more male than female Undeclared majors.

Conclusions. As with the clusters, the evidence supported three distinct classes. Although the classes were not distinct on help-seeking threat, overall they exhibited unique patterns across the classification variables. Class differences on the validity variables strongly supported a distinct *Class 3*, which exhibited the least adaptive pattern of means – significantly higher on help-seeking avoidance and significantly lower on self-acceptance, conscientiousness, and openness than the other classes. Additionally, class differences on gender and major exhibited noteworthy patterns, particularly when considered together.

Further research should explore whether similar classes are supported on other independent samples, and whether high proportions of Undeclared majors continue to be represented in a class that exhibits a less adaptive pattern of means. If so, more research is needed on why this is the case. Furthermore, additional research is needed on why

help-seeking threat played such a negligible role in distinguishing among the classes, particularly given what a comparatively large role this variable played in distinguishing the clusters in the cluster analysis solution.

What Do These Profiles Reveal?

Differences between cluster analysis and mixture modeling. Despite the fact that the aim of both cluster analysis and mixture modeling is to create groups of objects (persons) based on their responses to a set of variables, both analyses employ quite different methodologies. Cluster analysis is non-inferential and sample specific. Clusters are identified based solely on persons' similarity to one another on the clustering variables (Milligan & Hirtle, 2012) – that is, how close they are to one another in multivariate space (Everitt et al., 2011). In contrast, mixture modeling is a model-based procedure. It imposes a particular structure of means, variances, and covariances onto the classes and will only create the number of classes specified by the researcher (Bauer & Curran, 2004; Pastor & Gagné, 2013). Thus, the analyses' different approaches to creating groups would be expected to result in classification solution differences.

Final solution differences. These differences can be seen when examining the cluster and class solutions from the current study. Although there was a good deal of overlap in the cluster and class assignment (see Table 11, keeping in mind that the class variable is based on modal assignment), there was also considerable non-overlap, particularly when considering Cluster/Class 3. The overall ranking was largely the same between clusters and classes for all but two of the classification variables (WAV and HST). Most striking is the difference between the cluster and class solution on HST, as there were essentially no differences among the classes on HST. Thus, the classes were

more strongly differentiated from one another on the goal orientation variables (MAP, PAP, and PAV), whereas the clusters differed from one another across all the variables.

One possible explanation for this relates to Cluster/Class 3. Because Class 3 was much larger than both the other classes and Cluster 3, perhaps the larger size resulted in means that were closer to the total sample mean. In the cluster analysis, Clusters 1 and 2 were fairly similar on WAV, HST, and EHS; it was Cluster 3 that was clearly separated from the others. In the mixture modeling analysis, Class 3 was not very distinct from the other classes, resulting in classes whose means were lumped together on WAV, HST, and EHS.

The difference in cluster and class sizes begs the question of why the distribution of respondents across the clusters was so much more equal (n 's of 420, 340, and 471, respectively) than the distribution of respondents across the classes (n 's of 239, 184, and 808, respectively). The different algorithms used by cluster analysis versus mixture modeling are one likely reason for this. As already mentioned, mixture modeling imposes a structure on the data, such that the ultimate solution is the best one *based on the specified parameterization*, given the data. The parameterization specified here may have forced the uneven class sizes in order to fit the requirements (i.e., constrained between-class covariances, freely estimated within-class covariances, and freely estimated within- and between-class variances). In contrast, cluster analysis creates groups based on distance between variables. This could explain the discrepancy in the sizes of the mixture modeling solutions versus the cluster analysis solution.

Validity evidence. The validity variable analyses provided further evidence that the clusters and classes may be qualitatively different. The continuous validity variables'

patterns were similar in the class and clustering solution – Cluster/Class 3 was significantly different (in the expected direction) from the other clusters/classes, and Clusters/Classes 1 and 2 were not significantly different from one another on HSA, self-acceptance, or conscientiousness. However, unlike the cluster analysis solution, Class 1 reported significantly lower mean openness than Class 2, suggesting a possible qualitative difference between Clusters 1 and 2 and Classes 1 and 2. This idea was supported by the major distribution across the clusters and classes. Cluster 1 and Class 1 both included more Nursing majors than expected by chance; however, Cluster 2 consisted of more Education majors than expected by chance whereas Class 2 included more Arts and Humanities majors. This major distribution across classes makes sense when examining the wording of some items on the openness subscale. For example, students responded to openness items such as, “I see myself as someone who values artistic and aesthetic experiences” and “I see myself as someone who has few artistic interests (reverse-worded)”. Given the wording of the openness items, it is not surprising that Class 2 (e.g., Arts and Humanities) students scored significantly higher on openness than Class 1 students (e.g., Nursing). Cluster 3 included more Business/Economics majors than expected; this was not replicated in Class 3, which instead consisted of more Undeclared majors than expected by chance. These different proportions of majors across the classes suggests a difference in the qualitative composition of the two grouping solutions.

So which is “better” – mixture modeling or cluster analysis? As with many questions asking whether one thing is “better” than another, the answer is that it depends. As has been discussed, the different algorithms used to group persons may result in

similar, but still qualitatively distinct, clusters versus classes. Therefore, which analysis is best for a given study may depend on one's research questions. If a researcher is interested in sample data only and is opting to take a highly exploratory approach, cluster analysis may be a good choice. If, however, a researcher wants to make inferences to a population, has a strong, theory-based hypothesis about the structure of that population, can identify an appropriate parameterization, and has the appropriate software and skills required, mixture modeling might be the best approach. Mixture modeling is also an exploratory approach in that different numbers of classes and/or different parameterizations are typically specified. However, for a researcher who has absolutely no idea where to begin, the myriad of possible options available in mixture modeling may be unnecessarily complex and a hierarchical cluster analysis a more practical choice.

Student success. As indicated in the regression analysis, cluster assignment significantly predicted GPA, with the GPA of Clusters 1 and 2 (the adaptive and moderately adaptive clusters) being significantly higher than that of Cluster 3 (the cluster with the least adaptive profile). Furthermore, adding class assignment explained significantly more variance in GPA. Examination of the *b*-values indicated that this increased explanatory power was contributed entirely by Class 2, which, as described above, was the class with the moderately adaptive profile consisting of a proportionately large number of Arts and Humanities majors. This class's GPA was higher than both Class 1 and Class 3 (the class with the least adaptive profile).

However, these findings – though statistically significant – were not practically significant; overall, the model only explained 1.6% of the variance in GPA. The largest effect size seen in Table 13 is the sr^2 for Class 2, and that was only .01 when controlling

for the other predictors in the model (i.e., the clusters and Class 1). According to Cohen's (1988) benchmarks, this is a small effect – and in practical terms, it suggests that Class 2 only explained 1% of the variance in GPA. Thus, although it is tempting to interpret the findings as supporting the idea that the mixture modeling classes explain a significant amount of variance in GPA above and beyond what is explained by the clusters, such an interpretation may not be warranted given the miniscule effect sizes.

As an additional note, the comparison of non-nested regression models (predicting GPA from the clusters, and predicting GPA from the classes) indicated that the cluster and class models did not significantly differ in their explanatory power. This is most likely because the correlation between the two models ($r = .225$) was high – that is, the clusters and classes shared overlapping variance in the prediction of GPA. This result may speak to the question of which analysis is “better”. From a practical standpoint (i.e., ability to predict GPA in this sample), the answer to this question could thus be “neither”.

However, the Cohen's d and r^2 comparisons should also be considered. When modal assignment was used, the effect size differences in GPA were still small, like they were in the regression analysis. But when fractional class membership was allowed, the effect sizes were larger. Not only do these results support the idea that modal assignment should not be considered best practice, they also suggest that there may in fact be a difference in the clusters' and classes' ability to explain variance in GPA.

Implications, Limitations, and Future Research

One thing that is important to keep in mind when interpreting classification analyses is that the groupings should not be taken as absolute. That is, although they have been identified using statistical algorithms, they do not necessarily “actually” exist in the

population. Because cluster analysis is non-inferential, users cannot make this claim at all; but even mixture modeling, which (when using a direct approach) does allow the assumption that the classes actually exist in the population, should be interpreted cautiously. As already discussed, mixture modeling imposes a certain parameterization on the classes, which in turn produces a solution based on that parameterization. However, if the parameterization is misspecified, the classes will be misspecified as well. Additionally, a mixture model will output the number of classes requested, even if there are actually no classes in the population. Thus, though helpful, groupings that are identified via classification analyses should be interpreted while taking care to not make too strong a statement about their actual existence in the population.

An additional consideration with any classification analysis is the choice of variables. As outlined in the literature review, there were clear theoretical reasons for choosing the grouping and validity that were selected for this study. However, although the clusters and classes significantly predicted GPA, their predictive ability was weak (as per the small effect size). Had different variables been selected, the profiles' explanatory ability may have been greater. Thus, we should not give up on the idea of finding a set of variables that, when used in cluster analysis or mixture modeling, are able to predict GPA. As an exploratory study, this was merely the first step in finding the optimal set of variables and future research in this area is warranted. As an additional area for future research, academic outcomes other than GPA should be investigated. Perhaps the clusters and classes identified here would explain a practically significant amount of variance in some other outcome.

Similarly, researchers may want to consider validity variables other than those included in this study. Despite qualitatively distinct clusters and classes, none of the continuous variables distinguished between Clusters 1 and 2, and only three of the four continuous variables distinguished between Classes 1 and 2. Thus, we did not receive as much information as we could have about what makes these clusters and classes distinct from one another. Selecting other variables may shed more light on these distinctions.

Although the achievement goal orientation variables (Pastor et al., 2007) and help-seeking variables (Finney et al., 2014; White & Bembenuddy 2013) have been examined via person-centered analyses before, this is the first study that has combined them to identify profiles of students. Despite the fact that the regression analyses indicated that neither the clusters nor the classes practically significantly predicted GPA, Cohen's *d* comparisons suggested that GPA did practically significantly differ across fractionally-assigned classes. Classes containing more academically successful students (i.e., Classes 1 and 2) were characterized by high levels of mastery approach and low levels of work avoidance and executive help-seeking. In contrast, the class with the lowest GPA (i.e., Class 3) was characterized by low levels of mastery approach and relatively high levels of work avoidance and executive help-seeking. Educators should thus consider creating learning environments that foster adaptive learning strategies. For example, classrooms that promote a mastery approach orientation via cooperative work and informative feedback may assist students in the development of adaptive strategies, as could the encouragement of adaptive forms of help-seeking. Educators should also be on the lookout for students exhibiting maladaptive patterns of these characteristics, which could provide opportunities for intervention early on. GPA is only one aspect of

academic achievement and success, but the development of adaptive strategies could assist students in other academic areas, as well.

Conclusion

Any researcher who would like to adhere to the principles of Marsh and Hau's (2007) methodological synergy – the combination of substantive research and sound methodological practices – must consider the utility of person-centered techniques. Certainly, these analyses are not appropriate for every study; they may even need to be used alongside other, variable-centered techniques. However, it is the wise researcher who carefully considers his or her research questions before selecting an analysis, as opposed to simply selecting a technique that is most familiar.

This paper has not only described how to go about conducting two useful person-centered analyses, but has also demonstrated their similarities and differences using real data. Although the clusters and classes did not practically significantly predict GPA, the ease with which multiple patterns of means could be observed was a testament to the utility of classification analyses in understanding data. Despite being qualitatively and statistically different in many ways, each analysis has advantages and disadvantages that should be considered prior to selecting one or the other. Overall, however, it is our hope that researchers will consider person-centered analyses, where appropriate, for their own research in the future.

Sometimes, persons really can tell us more than just variables.

Tables

Table 1

Example of using agglomeration coefficients as a stopping rule.

Stage	Coefficients	Difference	# of Clusters	
...	
88	353.249	25.47386	8	
89	379.970	26.72107	7	
90	409.591	29.62002	6	
91	447.194	37.60322	5	← Large jump
92	502.173	54.97894	4	← Larger (more substantial?) jump
93	561.934	59.76140	3	
94	752.000	190.06592	2	

Table 2

Demographic Information for Participants

	<i>n</i> (%)
Gender	
Female	780 (63.4%)
Male	451 (36.6%)
Ethnicity	
American Indian	1 (.1%)
Asian	62 (5.0%)
Black	41 (3.3%)
Hispanic	26 (2.1%)
Pacific Islander	5 (.4%)
White	1034 (84.0%)
Not Specified	62 (5.0%)
Total <i>n</i>	1231
Age: Mean (<i>SD</i>)	18.43 (.40)

Table 3

Chi-square Results: Gender by Major

		Business/ Economics	Social Sciences	Arts & Humanities	Health Sciences	STEM majors	Education	Nursing	Undeclared
Female	Observed	97	92	77	127	93	71	65	158
	Expected	129.9	81.1	72.9	111.5	128.6	46.3	41.2	168.5
	Stand. Resid.	-2.9	1.2	.5	1.5	-3.1	3.6	3.7	-.8
Male	Observed	108	36	38	49	110	2	0	108
	Expected	75.1	46.9	42.1	64.5	74.4	26.7	23.8	97.5
	Stand. Resid.	3.8	-1.6	-.6	-1.9	4.1	-4.8	-4.9	1.1

Note: $\chi^2(7) = 135.69, p < .001$

Table 4

Subscale Means and Intercorrelations: Classification (above the Line) and Validity (below the Line) Variables (n = 1231)

	MAP	PAP	PAV	WAV	HST	EHS	HSA	Consc.	Open.	S-Acc.
MAP	-									
PAP	.391	-								
PAV	.257	.403	-							
WAV	-.429	-.164	-.048	-						
HST	-.170	.038	.013	.201	-					
EHS	-.282	-.046	.067	.454	.313	-				
HSA	-.286	-.060	-.060	.257	.685	.332	-			
Consc.	.294	.165	.031	-.341	-.191	-.352	-.274	-		
Open.	.244	.065	-.014	-.199	-.045	-.235	-.110	.115	-	
S-Acc.	.205	.076	.031	-.167	-.323	-.221	-.283	.347	.155	-
Mean(SD)	17.34(2.9)	16.50(3.7)	15.16(3.7)	11.33(4.5)	7.52(3.4)	5.29(2.2)	6.52(2.9)	32.31(5.3)	35.34(6.3)	41.22(7.2)
α	.77	.88	.65	.77	.76	.70	.74	.78	.79	.84
Skew	-.64	-.93	-.50	.51	.75	.61	.80	-.09	-.13	-.60
Kurtosis	.09	1.01	.01	.35	.57	.60	.27	.03	-.11	.33

Note: MAP=mastery approach, PAP=performance approach, PAV=performance avoidance, WAV=work avoidance, HST=help-seeking threat, EHS=executive help-seeking, HSA=help-seeking avoidance, Consc.=conscientiousness, Open.=openness, S-Acc.=self-acceptance

Table 5
Agglomeration Coefficients - Last 10

Stage	Coefficients	Difference
1221	131.753	4.725
1222	136.478	4.966
1223	141.444	7.334
1224	148.778	8.039
1225	156.817	9.633
1226	166.450	13.689
1227	180.139	16.362
1228	196.501	32.618
1229	229.119	37.971
1230	267.090	4.725

Table 6
Means and SDs of Final Clustering Solution (n=1231)

	Mean (SD)		
	Cluster 1 n = 420	Cluster 2 n = 340	Cluster 3 n = 471
MAP	19.37 (1.72)	17.05 (2.6)	15.74 (2.74)
PAP	19.32 (1.97)	13.66 (3.94)	16.04 (2.96)
PAV	17.81 (2.61)	11.90 (3.2)	15.15 (2.92)
WAV	8.67 (3.58)	9.86 (3.25)	14.76 (3.77)
HST	6.69 (3.23)	5.95 (2.42)	9.38 (3.18)
EHS	4.32 (1.74)	4.12 (1.5)	6.99 (1.88)
HSA	5.50 (2.61)	5.54 (2.31)	8.14 (2.87)
Conscientiousness	34.01 (5.21)	33.14 (5.11)	29.93 (4.7)
Openness	36.37 (6.22)	36.03 (6.34)	33.93 (5.98)
Self-acceptance	42.83 (7.2)	42.35 (6.3)	38.98 (7.3)

Note: MAP=mastery approach, PAP=performance approach,
PAV=performance avoidance, WAV=work avoidance, HST=help-seeking
threat, EHS=executive help-seeking, HSA=help-seeking avoidance

Table 7
ANOVA Results for Continuous Validity Variables (Clusters)

	<i>F</i>	<i>p</i>	η^2	Cluster 1 vs. 2	Cluster 1 vs. 3	Cluster 2 vs. 3
Help-seeking avoidance	144.20	< .001	0.19	$p = .98$	$p < .001$	$p < .001$
Conscientiousness	82.32	< .001	0.12	$p = .06$	$p < .001$	$p < .001$
Openness	20.38	< .001	0.03	$p = .75$	$p < .001$	$p < .001$
Self-acceptance	36.60	< .001	0.06	$p = .65$	$p < .001$	$p < .001$

Note. Group comparison *p*-values are from Scheffe's post-hoc test. *N* = 1231

Table 8

Chi-square Results: Cluster (Cluster Analysis) and Class (Mixture Modeling) by Major

		Business/ Economics	Social Sciences	Arts & Humanities	Health Sciences	STEM majors	Education	Nursing	Undeclared
Cluster 1	Observed	62	36	43	67	83	17	36	76
	Expected	69.9	43.7	39.2	60.0	69.3	24.9	22.2	90.8
	Stand. Resid.	-.9	-1.2	.6	.9	1.7	-1.6	2.9	-1.5
Cluster 2	Observed	46	49	33	49	51	30	10	72
	Expected	56.6	35.4	31.8	48.6	56.1	20.2	18.0	73.5
	Stand. Resid.	-1.4	2.3	.2	.1	-.7	2.2	-1.9	-.2
Cluster 3	Observed	97	43	39	60	69	26	19	118
	Expected	78.4	49.0	44.0	67.3	77.7	27.9	24.9	101.8
	Stand. Resid.	2.1	-.9	-.8	-.9	-1.0	-.4	-1.2	1.6

Note. Cluster chi-square: $\chi^2 = 47.35$, $p < .001$

Table 9
Fit Indices for the Three Mixture Model Parameterizations

	AIC	BIC	SSABIC	LMR	Entropy	LL	# parameters
1-Class A	39342.38	39378.19	39355.96	NA	NA	-19664.19	7
1-Class B	38654.16	38715.55	38677.43	NA	NA	-19315.08	12
1-Class C	37514.00	37652.12	37566.35	NA	NA	-18730.00	27
2-Class A	38558.74	38635.47	38587.82	$p < .001$	0.768	-19264.37	15
2-Class B	37607.15	37735.04	37655.63	$p < .01$	0.932	-18778.58	25
2-Class C	37024.47	37229.09	37102.03	$p = .024$	0.907	-18529.45	40
3-Class A	38192.96	38310.62	38237.57	$p = .237$	0.661	-19073.48	23
3 Class B*	-	-	-	-	-	-	-
3 Class C	36771.73	37042.85	36874.50	$p = .010$	0.733	-18332.86	53
4-Class A	37903.52	38062.10	37963.63	$p = .029$	0.695	-18920.76	31
4 Class B*	-	-	-	-	-	-	-
4 Class C*	-	-	-	-	-	-	-
5-Class A	37667.36	37866.87	37742.99	$p = .068$	0.711	-18794.68	39
5-Class B*	-	-	-	-	-	-	-
5-Class C*	-	-	-	-	-	-	-

*LL did not replicate despite 1000 starts; models were not stable.

Table 10

Class Means by Classification and Validity (Auxiliary) Variables

Class Means based on Posterior Probabilities			
Measure	Class 1 <i>n</i> = 239	Class 2 <i>n</i> = 184	Class 3 <i>n</i> = 808
Mastery Approach	19.61	19.10	16.17
Performance Approach	19.94	14.87	15.84
Performance Avoidance	18.71	12.17	14.82
Work Avoidance	10.10	10.16	12.02
Help-seeking Threat	7.37	7.65	7.53
Executive Help-seeking	5.25	3.56	5.74
Help-seeking Avoidance	4.37 ^a	4.59 ^a	8.56 ^{b,c}
Conscientiousness	33.97 ^a	34.72 ^a	30.69 ^{b,c}
Openness	36.34 ^{a,b}	38.55 ^{a,c}	33.87 ^{b,c}
Self-Acceptance	42.96 ^a	42.83 ^a	40.17 ^{b,c}

^a = significantly ($p < .01$) different from Class 3, ^b = significantly ($p < .01$) different from Class 2, ^c = significantly ($p < .01$) different from Class 1, based on chi-square analyses

Table 11
Covariances and Variances by Class*

Class 1						
	MAP	PAP	PAV	WAV	HST	EHS
MAP	3.03					
PAP	2.36	2.45				
PAV	1.59	1.68	4.47			
WAV	-3.61	-2.10	-0.45	28.95		
HST	-1.36	-0.30	-0.06	2.49	14.87	
EHS	-0.64	-0.69	-0.26	3.56	2.24	6.03
Class 2						
	MAP	PAP	PAV	WAV	HST	EHS
MAP	3.32					
PAP	2.36	28.56				
PAV	1.59	1.68	18.15			
WAV	-3.61	-2.10	-0.45	23.12		
HST	-1.36	-0.30	-0.06	2.49	18.99	
EHS	-0.64	-0.69	-0.26	3.56	2.24	1.91
Class 3						
	MAP	PAP	PAV	WAV	HST	EHS
MAP	6.65					
PAP	2.36	8.94				
PAV	1.59	1.68	8.99			
WAV	-3.61	-2.10	-0.45	15.40		
HST	-1.36	-0.30	-0.06	2.49	8.17	
EHS	-0.64	-0.69	-0.26	3.56	2.24	4.09

* Variances are presented on the diagonal

Table 12
Classification Table: Cluster by Class

	Cluster 1	Cluster 2	Cluster 3	Total
Class 1	193	0	46	239
	80.8%	0.0%	19.2%	100%
Class 2	44	111	29	184
	23.9%	60.3%	15.8%	100%
Class 3	183	229	396	808
	22.6%	28.3%	49.0%	100%
Total	420	340	471	1231
	34.1%	27.6%	38.3%	100%

Table 13
*Regression Values for the Prediction of Spring GPA from Cluster and Class
 (Cluster/Class 3 as Comparison Group)*

Step and Predictor	R^2	95% CI of R^2	R^2 Change	b	95% CI of b	sr^2
Step 1	.005*	.000, .015	.005*			
Cluster 1				.095*	.012, .179	.004
Cluster 2				.090*	.001, .180	.003
Step 2	.016**	.003, .030	.011**			
Cluster 1				.090	-.001, .182	.003
Cluster 2				.039	-.054, .132	.001
Class 1				-.009	-.113, .095	.000
Class 2				.191**	.085, .297	.010
Step 3 [†]	.017	.002, .029	.001			
Cluster 1				.099	-.012, .210	.003
Cluster 2				.053	-.051, .157	.001
Class 1				.060	-.133, .253	.000
Class 2				.175	-.063, .414	.002
Cluster1 x Class1 Interaction				-.087	-.319, .145	.000
Cluster1 x Class2 Interaction				.070	-.246, .386	.000
Cluster2 x Class2 Interaction				-.010	-.288, .269	.000

* $p < .05$ ** $p < .01$

[†]The other 7 interaction variables dropped out of the analysis because they did not contribute to the model (b and $sr^2 = 0$).

Table 14
*Regression Values for the Prediction of Spring GPA from Cluster and Class
 (Cluster/Class 2 as Comparison Group)*

Step and Predictor	R^2	95% CI of R^2	R^2 Change	b	95% CI of b	sr^2
Step 1	.005*	.000, .015	.005*			
Cluster 1				.005	-.086, .096	.000
Cluster 3				-.090*	-.180, -.001	.003
Step 2	.016**	.003, .030	.011**			
Cluster 1				.052	-.052, .155	.001
Cluster 3				-.039	-.132, .054	.001
Class 1				-.200*	-.336, -.064	.007
Class 3				-.191**	-.297, -.085	.010
Step 3 [†]	.017	.002, .029	.001			
Cluster 1				.125	-.096, .347	.001
Cluster 3				-.053	-.157, .051	.001
Class 1				-.272*	-.479, -.065	.005
Class 3				-.166*	-.311, -.021	.004
Cluster1 x Class3 Interaction				-.080	-.333, .174	.000
Cluster3 x Class1 Interaction				.167	-.151, .485	.001
Cluster3 x Class2 Interaction				.010	-.269, .288	.000

* $p < .05$ ** $p < .01$

[†]The other 7 interaction variables dropped out of the analysis because they were made up entirely of zeroes.

Table 15
*Cohen's d Comparison of GPA Means across Classes (by
 Assignment Type) and Clusters*

		d	χ^2*
1 vs. 2	Modally-assigned class	.28	-
	Fractional class	.45	23.11**
	Cluster	.01	-
1 vs. 3	Modally-assigned class	.05	-
	Fractional class	.03	0.113
	Cluster	.06	-
2 vs. 3	Modally-assigned class	.33	-
	Fractional class	.44	41.51**
	Cluster	.05	-

*Chi-square comparison from output entering GPA as an auxiliary variable ** $p < .001$

Figures

Figure 1. Illustration of how structure can be imposed on data where no structure exists.

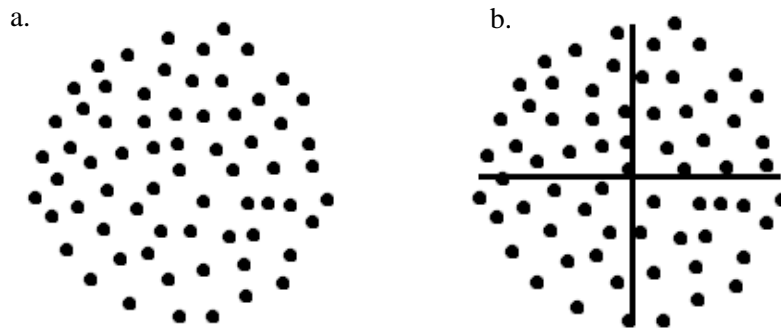


Figure 2. Illustration of the issues with using correlation as a measure of similarity.

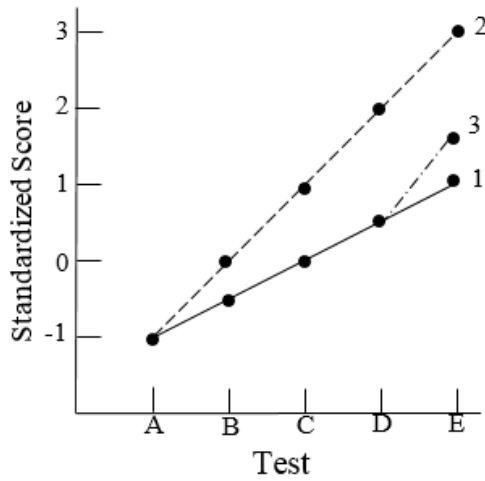


Figure 3. Visual representation of the concept of Euclidean distance.

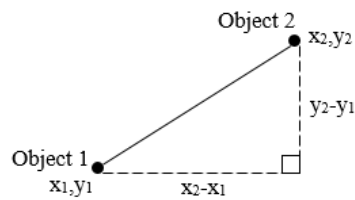


Figure 4. Possible student profiles resulting from cluster analysis or mixture modeling, utilizing the variables of study.

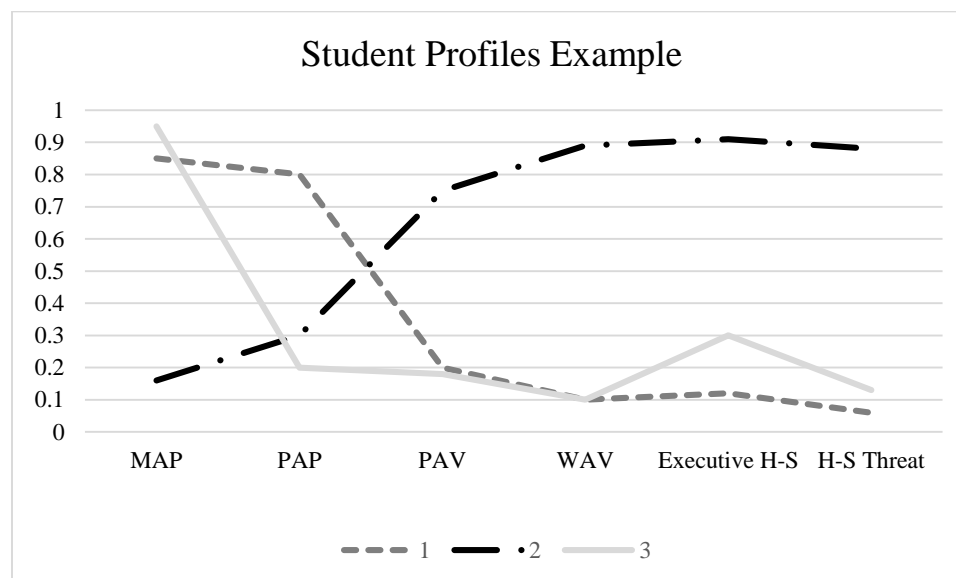


Figure 5. Z-score means by cluster for the three-cluster hierarchical agglomerative cluster analysis solution.

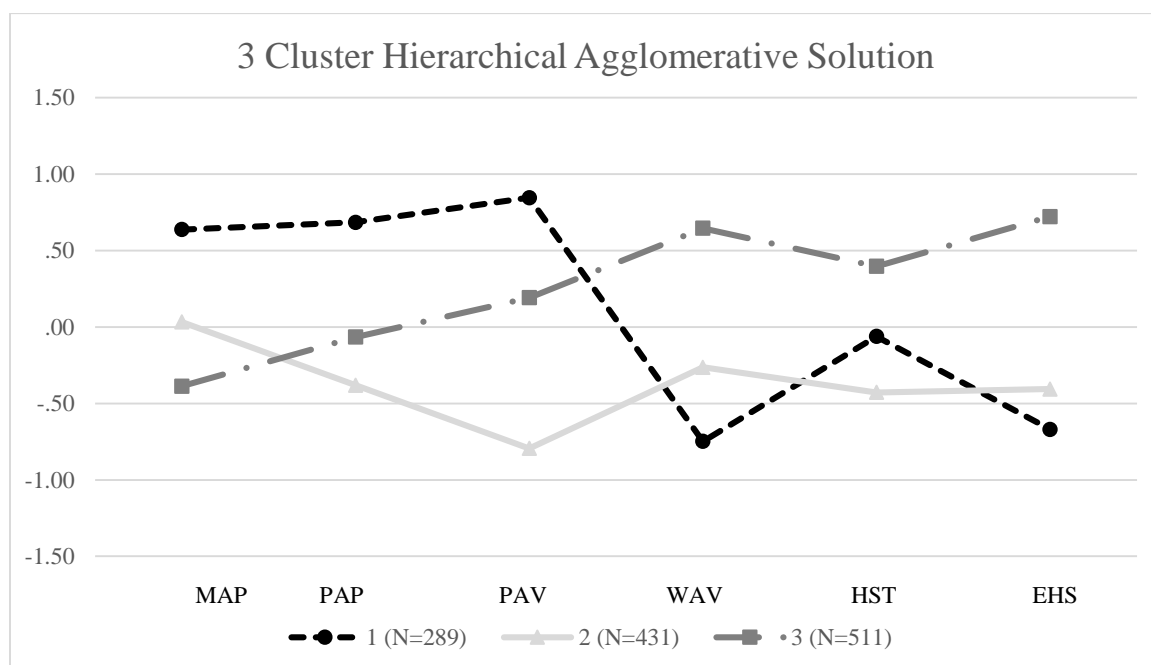


Figure 6. Z-score means by cluster for the final three-cluster k -means cluster analysis solution.

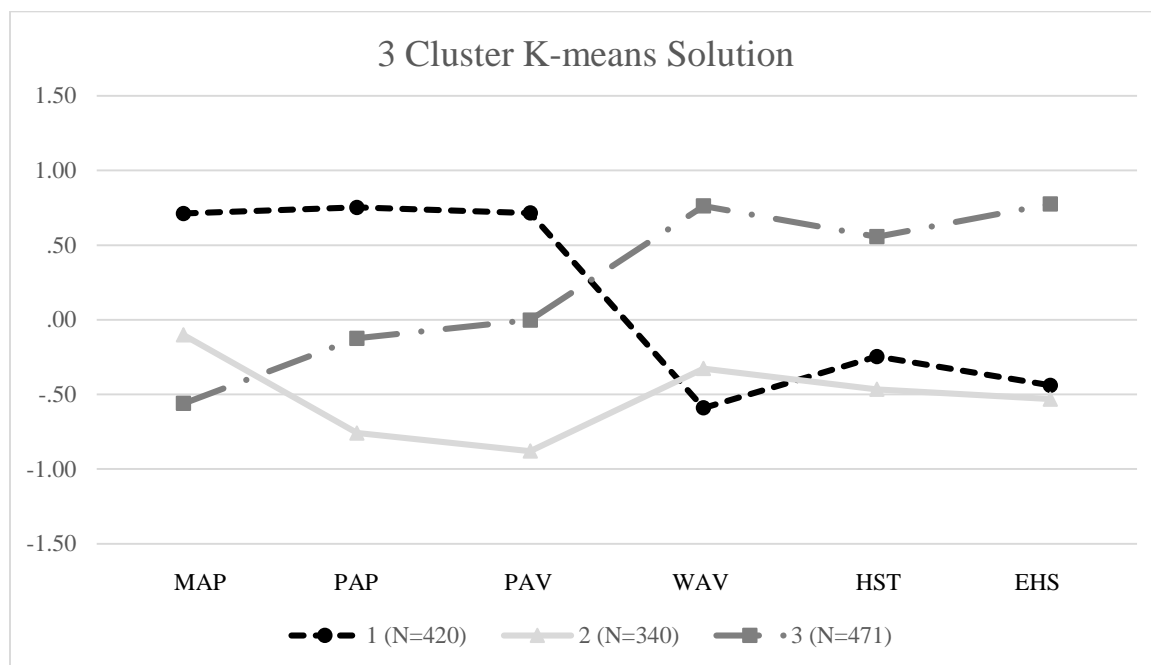
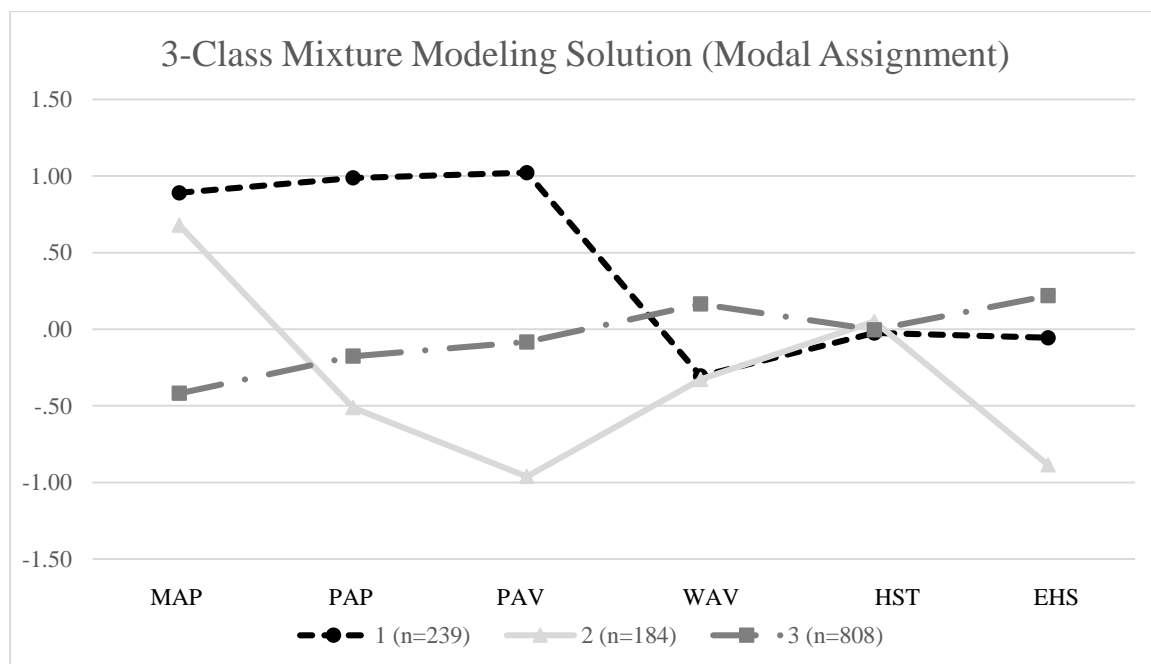


Figure 7. Z-score means by class for the final three-class mixture modeling solution (modal assignment).



Appendix A

Description of Affective and Attitudinal Measures Completed by Students at Both Time Points

Subtest	Subscales	Sample Item	Scale Range
Achievement Goal Questionnaire	Mastery-Approach (3 items)	“My aim is to completely master the material in my courses this semester.”	1 (not at all true of me) to 7(very true of me)
	Performance-Approach (3 items)	“My aim this semester is to perform well relative to other students.”	
	Performance-Avoidance (3 items)	“My aim to avoid doing worse than other students.”	
	Work-Avoidance (4 items)	“I want to do as little work as possible this semester.”	
Help-Seeking Scale	Executive Help-Seeking (2 items)	“Getting help in this class would be a way of avoiding doing some of the work.”	1 (strongly disagree) to 8 (strongly agree)
	Help-Seeking Threat (3 items)	“I would feel like a failure if I needed help in this class.”	
	Help-Seeking Avoidance (3 items)	“I would rather do worse on an assignment I couldn’t finish than ask for help.”	
Psychological Well-being Scale	Self-Acceptance (9 items)	“In general, I feel confident and positive about myself.”	1 (strongly disagree) to 6 (strongly agree)
Big Five Inventory	Conscientiousness (9 items)	“I see myself as someone who does a thorough job.”	1 (disagree strongly) to 5 (agree strongly)
	Openness (10 items)	“I see myself as someone who has an active imagination.”	

References

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrov & F. Csake (Eds.). *Second international symposium on information theory* (pp. 267–281). Budapest: Akademiai Kiado.
- Ames, C. (1984). Achievement attributions and self-instructions under competitive and individualistic goal structures. *Journal of Educational Psychology*, 76(3), 478-487. doi:10.1037/0022-0663.76.3.478
- Anderberg, M.R. (1973). *Cluster analysis for applications*. New York, NY: Academic Press, Inc.
- Asparouhov, T., & Muthén, B. (2013). Auxiliary variables in mixture modeling: 3-step approaches using Mplus. *Mplus web notes*, 15, 1-24.
- Baker, F.B. (1974). Stability of two hierarchical grouping techniques case 1: Sensitivity to data errors. *Journal of the American Statistical Association*, 69(346), 440-445.
- Baker, F.B., & Hubert, L.J. (1975). Measuring the power of hierarchical cluster analysis. *Journal of the American Statistical Association*, 70, 31-38.
- Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44, 1-26.
- Barron, K.E., & Harackiewicz, J.M. (2001). Achievement goals and optimal motivation: Testing multiple goal models. *Journal of Personality and Social Psychology*, 80(5), 706-722.
- Barron, K.E., & Harackiewicz, J.M. (2003). Revisiting the benefits of performance-approach goals in the college classroom: Exploring the role of goals in advanced college courses. *International Journal of Educational Research*, 39, 357-374.

- Barry, C. L., Horst, S. J., Brown, A. R., Finney, S. J., & Kopp, J. P. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International Journal of Testing*, 10, 342-363.
- Bauer, D.J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research*, 42(4), 757-786.
- Bauer, D.J., & Curran, P.J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, 9(1), 3-29. doi: 10.1037/1082-989X.9.1.3
- Bauer, D.J., & Shanahan, M.J. (2007). Modeling complex interactions: Person-centered and variable-centered approaches. In Little, T.D., Bovaird, J.A. & Card, N.A. (Eds.). *Modeling ecological and contextual effects in longitudinal studies of human development* (pp. 255-283). Mahwah, NJ: LEA.
- Benet-Martínez, V., & John, O.P. (1998). Los Cinco Grandes across cultures and ethnic groups: Multitrait-multimethod analyses of the Big Five in Spanish and English. *Journal of Personality and Social Psychology*, 75(3), 729-750. doi: 10.1037/0022-3514.75.3.729
- Bergman, L.R., & Magnusson, D. (1997). A person-oriented approach in research on developmental psychopathology. *Development and Psychopathology*, 9, 291-319.
- Blashfield, R.K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, 83(3), 377-388.
- Bozdogan, H. (1987). Model selection and Akaike's Information Criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52, 345-370.

- Breckenridge, J.N. (1989). Replicating cluster analysis: Method, consistency, and validity. *Multivariate Behavioral Research*, 24(2), 147-161. doi: 10.1207/s15327906mbr2402_1
- Brophy, J. (1983). Conceptualizing student motivation. *Educational Psychologist*, 18(3), 200-215.
- Caliński, R.B., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1-27.
- Chemers, M.M., Hu, L., & Garcia, B.F. (2001). Academic self-efficacy and first-year college student performance and adjustment. *Journal of Educational Psychology*, 93(1), 55-64. doi: 10.1037//0022-0663.93.1.55
- Clark, S.L. (2010). *Mixture modeling with behavioral data* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3405665)
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J. (1992). A power primer. *Quantitative Methods in Psychology*, 112(1), 155-159.
- Coleman, J. (1986). Social theory, social research, and a theory of action. *American Journal of Sociology*, 91(6), 1309-1335.
- Conard, M.A. (2006). Aptitude is not enough: How personality and behavior predict academic performance. *Journal of Research in Personality*, 40, 339-346.
- Davidson, C.N. (2008). Humanities 2.0: Promise, perils, predictions. *PMLA*, 123(3), 707-717.
- de Raad, B., & Schoewenburg, H.C. (1996). Personality in learning and education: A review. *European Journal of Personality*, 10, 303-336.

- DeBerard, M. S., Spielmans, G.I., & Julka, D.C. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal*, 38(1), 66-80.
- DiStefano, C., & Kamphaus, R.W. (2006). Investigating subtypes of child development: A comparison of cluster analysis and latent class cluster analysis in typology creation. *Education and Psychological Measurement*, 66(5), 778-794. doi: 10.1177/0013164405284033
- Duda, R.O., & Hart, P.E. (1973). *Pattern classification and scene analysis*. New York: Wiley.
- Dudek, M.W.A. (2014). clusterSim: Searching for optimal clustering procedure for a data set (version 0.43-5) [Computer software]. Retrieved from <http://CRAN.R-project.org/package=clusterSim>
- Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist*, 41(10), 1040-1048. Retrieved from <http://www.nisdx.org/cms/lib/TX21000351/Centricity/Domain/21/j%20carlisle/Motivational%20Processes.pdf>
- Edwards, A.W.F., & Cavalli-Sforza, L.L. (1965). A method for cluster analysis. *Biometrics*, 21(2), 362-375.
- Elliot, A.J., & McGregor, H.A. (2001). A 2x2 achievement goal framework. *Journal of Personality and Social Psychology*, 80(3), 501-519. doi:10.1037//0022-3514.80.3.501

- Elliot, A.J., McGregor, A.H., & Gable, S. (1999). Achievement goals, study strategies, and exam performance: A mediational analysis. *Journal of Educational Psychology, 91*(3), 549-563.
- Enders, C.K. (2005). Maximum likelihood estimate. In B.S. Everitt & D.C. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (1164-1170). Chichester, John Wiley & Sons.
- Everitt, B.S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). West Sussex, UK: John Wiley & Sons.
- Finch, W. H., & Bronk, K. C. (2011). Conducting confirmatory latent class analysis using Mplus. *Structural Equation Modeling: A Multidisciplinary Journal, 18*, 132–151.
- Finney, S.J., Barry, C.L., Horst, S.J., & Johnston, M.M. (2014). *Are there qualitatively distinct academic help-seeking types? An application of mixture modeling*. Manuscript submitted for publication.
- Finney, S.J., Pieper, S.L., & Barron, K.E. (2004). Examining the psychometric properties of the Achievement Goal Questionnaire in general academic context. *Educational and Psychological Measurement, 64*(2), 265-382. doi: 10.1177/0013164403258465
- Fleiss, J.L., & Zubin, J. (1969). On the methods and theory of clustering. *Multivariate Behavioral Research, 4*(2), 235-250.
- Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*, 49-66.
- Hartigan, J. (1975). *Clustering algorithms*. New York: Wiley.

- Hair, J.F., Anderson, R.E., Tatham, R.L., & Black, W. (1998). *Multivariate data analysis* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Harackiewicz, J.M., Barron, K.E., Tauer, J.M., Carter, S.M., & Elliot, A.J. (2000). Short-term and long-term consequences of achievement goals: Predicting interest and performance over time. *Journal of Educational Psychology*, 92, 316-330.
- Henson, J. M., Reise, S. P., & Kim, K. H. (2007). Detecting mixtures from structural model differences using latent variable mixture modeling: A comparison of relative model fit statistics. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 202–226.
- Hipp, J.R., & Bauer, D.J. (2006). Local solutions in the estimation of growth mixture models. *Psychological Methods*, 11(1), 36-53. doi: 10.1037/1082-989X.11.1.36
- Huq, M., Rabman, M.M., & Mahmud, S.H. (1986). Role of neuroticism, psychoticism, and extraversion in academic achievement. *Asian Journal of Psychology and Education*, 17(2), 1-6.
- Jedidi, K., Jagpal, H. S., & DeSarbo, W. S. (1997). Finite mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science*, 16, 39–59.
- John, O.P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L.A. Pervin and O.P. John (Eds.). *Handbook of personality: Theory and research* (pp. 102-138). New York, NY: Guilford.
- John, O.P., Donahue, E.M., & Kentle, R.L. (1991). *The Big Five Inventory*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

- Johnson, S.C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241-254.
- Karabenick, S.A. (2003). Seeking help in large college classes: A person-centered approach. *Contemporary Educational Psychology*, 28, 37-58.
- Karabenick, S.A., & Dembo, M.H. (2011). Understanding and facilitating self-regulated help seeking. *New Directions for Teaching and Learning*, 126, 33-43.
- Karabenick, S.A., & Knapp, J.R. (1991). Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. *Journal of Educational Psychology*, 83(2), 221-230.
- Kuiper, F.K., & Fisher, L. (1975). A Monte Carlo comparison of six clustering procedures. *Biometrics*, 31(3), 777-783.
- Lanza, S.T., Tan, X., & Bray, B.C. (2013). Latent class analysis with distal outcomes: A flexible model-based approach. *Structural Equation Modeling*, 20, 1-26.
- Laursen, B., & Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly*, 52(3), 377-389.
- Linnenbrink, E.A., & Pintrich, P.R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, 31(3), 313-327.
- Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, 88, 767-778.
- Lorr, M. (1983). *Cluster analysis for the social sciences*. London: Jossey-Bass Inc.
- Lubke, G. (2010). Latent variable mixture models. In G.R. Hancock and R.O. Mueller, R. O. (Eds.). *The reviewer's guide to quantitative methods in the social sciences* (pp. 209-219). New York, NY: Routledge.

- MacCallum, R.C., Zhang, S., Preacher, K.J., & Rucker, D.D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, 7, 19-40.
- Magidson, J., & Vermunt, J.K. (2002). Latent class models for clustering: A comparison with k-means. *Canadian Journal of Marketing Research*, 20(1), 2002.
- Magnusson, D. (1998). The logic and implications of a person-oriented approach. In R.B. Cairns, L.R. Bergman, & J. Kagan (Eds.), *Methods and models for studying the individual* (pp. 33-63). Thousand Oaks, CA: Sage.
- Marsh, H.W., & Hau, K.T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32, 151-170.
- Marsh, H.W., Lüdtke, O., Trautwein, U., & Morin, A.J.S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling*, 16(2), 191-225. doi: 10.1080/10705510902751010
- McCullough, M.E., Bellah, C.G., Kilpatrick, S.D., & Johnson, J.L. (2001). Vengefulness: Relationships with forgiveness, rumination, well-being, and the Big Five. *Personality and Social Psychology Bulletin*, 27, 601-610.
- McIntyre, R. M., & Blashfield, R.K. (1980). A nearest-centroid technique for evaluating the minimum-variance clustering procedure. *Multivariate Behavioral Research*, 15(2), 225-238. doi: 10.1207/s15327906mbr1502_7
- McLachlan, G.J., & Peel, D. (2000). *Finite mixture models*. New York, NY: Wiley.
- Meehl, P. E. (1992). Factors and taxa, traits and types, differences of degree and differences in kind. *Journal of Personality*, 60(1), 117-174.

- Milligan, G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45(3), 325-342.
- Milligan, G.W. (1996). Clustering validation: Results and implications for applied analyses. In P. Arabie, L.J. Hubert, & G. De Soete (Eds.). *Clustering and Classification* (pp. 341-379). River Edge, NJ: World Scientific Publications.
- Milligan, G.W., & Cooper, M.C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159-179.
- Milligan, G.W., & Cooper, M.C. (1987). Methodology review: Clustering methods. *Applied Psychological Measurement*, 11, 329-354. doi: 10.1177/014662168701100401
- Milligan, G.W., & Cooper, M.C. (1988). A study of standardization of variables in cluster analysis. *Journal of Classification*, 5(2), 181-204.
- Milligan, G.W., & Hirtle, S.C. (2012). Clustering and classification methods. In I.B. Weiner, J.A. Schinka, and W.F. Velicer (Eds.). *Handbook of Psychology: Research Methods in Psychology* (2nd ed.). Charlottesville, VA: John Wiley & Sons.
- Mojena, R. (1977). Hierarchical grouping methods and stopping rules: An evaluation. *The Computer Journal*, 20, 359-363.
- Mooney, S.P., Sherman, M.F., & LoPresto, C.T. (1991). Academic locus of control, self-esteem, and perceived distance from home as predictors of college adjustment. *Journal of Counseling and Development*, 69, 445-448.
- Nylund, K. L., Asparouhov, T., & Muthen, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo

- simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 535–569.
- Pastor, D.A. (2010). Cluster Analysis. In G.R. Hancock and R.O. Mueller, R. O. (Eds.). *The reviewer's guide to quantitative methods in the social sciences*. New York, NY: Routledge.
- Pastor, D.A., Barron, K.E., Miller, B.J., & Davis, S.L. (2007). A latent profile analysis of college students' achievement goal orientation. *Contemporary Educational Psychology*, 32, 8-47. doi: 10.1016/j.cedpsych.2006.10.003
- Pastor, D.A., & Gagné, P. (2013). Mean and covariance structure mixture models. In G.R. Hancock and R.O. Mueller, R. O. (Eds.). *Structural equation modeling: A second course* (2nd ed.). Charlotte, NC: Information Age Publishing Inc.
- Petersen, I., Louw, J., & Dumont, K. (2008). Adjustment to university and academic performance among disadvantaged students in South Africa. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 29(1), 99-115. doi: 10.1080/01443410802521066
- Pieper, S.L. (2003). *Refining and Extending the 2x2 Achievement Goal Framework: Another Look at Work Avoidance*. (Unpublished doctoral dissertation). James Madison University, Harrisonburg, VA.
- Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, 135(2), 322-338.
- Pyburn, E.M., Horst, S.J., & Erbacher, M. (October 2014). *International student success: An application of cluster analysis to predict GPA*. Paper presented at the annual conference of the Northeastern Educational Research Association, Trumbull, CT.

- R Core Team (2014). R: A language and environment for statistical computing (version 3.1.1) [Computer software]. Vienna: retrieved from <http://www.R-project.org>
- Raykowsky, D.A., & Lance, G.N. (1978). A criterion for determining the number of groups in a classification. *Australian Computer Journal*, 10, 115-117.
- Richardson, M., Bon, R., & Abraham, C. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138(2), 353-387. doi: 10.1037/a0026838
- Robbins, S.B., Davis, H.L., Lauver, K., & Langley, R. (2004). Do psychosocial and study skill factors predict college outcomes? A meta-analysis. *Psychological Bulletin*, 130(2), 261-288. doi: 10.1037/0033-2909.130.2.261
- Roussel, P., Elliot, A.J., & Feltman, R. (2011). The influence of achievement goals and social goals on help-seeking from peers in an academic context. *Learning and Instruction*, 21, 394-402.
- Ryff, C.D. (1989). Happiness is everything, or is it? Explorations of the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1069-1081. doi:10.1037/0022-3514.57.6.1069.
- Scheibler, D., & Schneider, W. (1985). Monte Carlo tests of the accuracy of cluster analysis algorithms: A comparison of hierarchical and nonhierarchical methods. *Multivariate Behavioral Research*, 20(3), 283-304. doi: 10.1207/s15327906mbr2003_4
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.

- Sclove, S. L. (1987). Application of model-selection criteria to some problems in multivariate analysis. *Psychometrika*, 52, 333–343.
- Shaver, P.R., & Brennan, K.A. (1992). Attachment styles and the “Big Five” personality traits: Their connections with each other and with romantic relationship outcomes. *Personality and Social Psychology Bulletin*, 19, 536-546.
- Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, 87(2), 245-251.
- Steinley, D. (2003). Local optima in *k*-means clustering: What you don’t know may hurt you. *Psychological Methods*, 8(3), 294-304.
- Steinley, D. (2004). Standardizing variables in k-means clustering. *Classification, clustering, and data mining applications: Proceedings of the meeting of the International Federation of Classification Societies*. Chicago, IL.
- Steinley, D. & Brusco, M.J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, 16(1), 63-79. doi: 0.1037/a0022673
- Strahan, E.Y. (2003). The effects of social anxiety and social skills on academic performance. *Personality and Individual Differences*, 34, 347-366. Retrieved from [http://dx.doi.org/10.1016/S0191-8869\(02\)00049-1](http://dx.doi.org/10.1016/S0191-8869(02)00049-1)
- Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics*. Boston: Pearson.
- Tan, P., Steinbach, M., & Kumar, V. (2006). Cluster Analysis: Basic Concepts and Algorithms. In *Introduction to data mining*. Boston, MA: Addison-Wesley.
- Tofghi, D., & Enders, C. K. (2008). Identifying the correct number of classes in growth mixture models. In G. R. Hancock & K. M. Samuelsen (Eds.), *Advances in latent*

- variable mixture models* (pp. 317–341). Greenwich, CT: Information Age Publishing, Inc.
- Trapmann, S., Hell, B., Hirn, J.W., & Schuler, H. (2007). Meta-analysis of the relationship between the Big Five and academic success at university. *Journal of Psychology, 215*(2), 132-151.
- Vermunt, J.K., & Magidson, J. (2002). Latent class cluster analysis. In J.A. Hagenaars and A.L. McCutcheon (Eds.). *Applied latent class analysis*. New York, NY: Cambridge University Press.
- Välimaa, J. (1998). Culture and identity in higher education research. *Higher Education, 36*, 119-138.
- Von Eye, A., & Bogat, A. (2006). Person-oriented and variable-oriented research: Concepts, results, and development. *Merrill-Palmer Quarterly, 52*(3), 390-420.
- Wang, C., Brown, C.H., & Bandeen-Roche, K. (2005). Residual diagnostics for growth mixture models: Examining the impact of a preventative intervention on multiple trajectories of aggressive behavior. *Journal of the American Statistical Association, 100*(471), 1054-1076.
- Wang, K.T., Heppner, P.P., Fu, C., Zhao, R., Li, F., & Chuang, C. (2012). Profiles of acculturative adjustment patterns among Chinese international students. *Journal of Counseling Psychology, 59*(3), 424-436. doi: 10.1037/a0028532
- Ward, J.H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236-244.

- White, M.C., & Bembenuddy, H. (2013). Not all avoidance help seekers are created equal: Individual differences in adaptive and executive help seeking. *SAGE Open*, 1-14. doi: 10.1177/2158244013484916
- Whiteman, S.D., & Loken, E. (2006). Comparing analytic techniques to classify dyadic relationships: An example using siblings. *Journal of Marriage and Family*, 68, 1370-1382.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594-604.
- Wintre, M.G., Diloura, B, Pancer, S.M., Pratt, M.W., Birnie-Lefcovitch, S., Polivy, J., & Adams, G. (2011). Academic achievement in first-year university: Who maintains their high school average? *Journal of Higher Education*, 62, 467-481. doi: 10.1007/s10734-010-9399-2
- Yang, C.C. (2006). Evaluating latent class analysis models in qualitative phenotype identification. *Computational Statistical & Data Analysis*, 50, 1090–1104.
- Zusho, A., Pintrich, P.R., & Cortina, K.S. (2005). Motives, goals, and adaptive patterns of performance in Asian-American and Anglo-American students. *Learning and Individual Differences*, 15, 141-158.